OM nucleic - nucleic search, using sw model

```
            GenCore version 4.5
     Copyright (c) 1993 - 2000  Compugen Ltd.
```

Run on: May 25, 2000, 23:47:05 ; Search time 550.28 Seconds
                                            (without alignments)
                                            -789.343 Million cell updates/sec

Title: US-09-099-898-1
Perfect score: 453
Sequence: 1 GTGTGCCGGATTGGTTAGC.........GGGGAGCTGGAAATAAACCT 453

Scoring table: IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                  Listing first 45 summaries

Database :   GenEmbl:*
                  1: gb_bal:*
                  2: gb_ba2:*
                  3: gb_om:*
                  4: gb_ov:*
                  5: gb_pat:*
                  6: gb_ph:*
                  7: gb_pl1:*
                  8: gb_pl2:*
                  9: gb_prl:*
                  10: gb_pr2:*
                  11: gb_pr3:*
                  12: gb_ro:*
                  13: gb_sts:*
                  14: gb_sy:*
                  15: gb_un:*
                  16: gb_vi:*
                  17: em_fun:*
                  18: em_hum1:*
                  19: em_hum2:*
                  20: em_in:*
                  21: em_om:*
                  22: em_or:*
                  23: em_ov:*
                  24: em_pat:*
                  25: em_ph:*
                  26: em_pl:*
                  27: em_ro:*
                  28: em_sts:*
                  29: em_sy:*
                  30: em_un:*
                  31: em_vi:*
                  32: gb_htg1:*
                  33: gb_htg2:*
                  34: gb_in1:*
                  35: gb_in2:*
                  36: gb_ba1:*
                  37: em_ba2:*
                  38: em_hum3:*
                  39: em_hum4:*
                  40: gb_pr4:*
                  41: gb_htg3:*
                  42: gb_htg4:*
                  43: gb_htg5:*
                  44: gb_htg6:*
                  45: gb_htg7:*
                  46: em_htg1:*
                  47: em_htg2:*
                  48: em_htg3:*
                  49: em_hum5:*
                  50: gb_pl3:*
                  51: gb_pr5:*
                  52: gb_htg8:*
                  53: gb_htg9:*
                  54: gb_htg10:*
                  55: gb_htg11:*
                  56: gb_htg12:*
                  57: gb_htg13:*
                  58: gb_htg14:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result NO. | | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|---|
| 1 | | 129 | 28.5 | 214575 | 45 | AC008763 | AC008763 Homo sapi |
| 2 | | 69.6 | 15.4 | 90442 | 57 | AC016932 | AC016932 Homo sapi |
| 3 | c | 45.8 | 10.1 | 90442 | 57 | AC016932 | AC016932 Homo sapi |
| 4 | | 43.8 | 9.7 | 981 | 12 | RNCKERIMP | X54806 R.norvegicu |
| 5 | | 43.4 | 9.6 | 41055 | 1 | SC6G4 | AL031317 Streptomy |
| 6 | | 42.8 | 9.4 | 207651 | 55 | AC009954 | AC009954 Homo sapi |
| 7 | c | 42.8 | 9.4 | 219565 | 53 | AC009974 | AC009974 Homo sapi |
| 8 | | 42.6 | 9.4 | 76968 | 54 | AC022648 | AC022648 Homo sapi |
| 9 | | 42.6 | 9.4 | 164520 | 44 | AC020738 | AC020738 Homo sapi |
| 10 | | 42.4 | 9.2 | 174707 | 43 | AC017082 | AC017082 Homo sapi |
| 11 | | 41.8 | 9.2 | 76199 | 55 | AC023395 | AC023395 Homo sapi |
| 12 | c | 41.8 | 9.2 | 81213 | 55 | AC023660 | AC023660 Homo sapi |
| 13 | | 41.6 | 9.2 | 40883 | 11 | AC005776 | AC005776 Homo sapi |
| 14 | c | 41.2 | 9.1 | 68523 | 55 | AC023214 | AC023214 Homo sapi |
| 15 | c | 41.2 | 9.1 | 75135 | 55 | AC005445 | AC005445 Drosophil |
| 16 | | 40.8 | 9.0 | 5180 | 12 | RNU48596 | U48596 Rattus norv |
| 17 | | 40.8 | 9.0 | 39441 | 11 | AC005565 | AC005565 Homo sapi |
| 18 | | 40.8 | 9.0 | 96079 | 56 | AC023362 | AC023362 Mus muscu |
| 19 | | 40.8 | 9.0 | 127778 | 44 | AC020735 | AC020735 Homo sapi |
| 20 | | 40.6 | 9.0 | 47852 | 1 | MTV023 | AL022022 Mycobacte |
| 21 | | 40.6 | 9.0 | 55482 | 54 | AC022663 | AC022663 Homo sapi |
| 22 | | 40.6 | 9.0 | 171480 | 44 | AC021165 | AC021165 Homo sapi |
| 23 | c | 40.6 | 9.0 | 196696 | 55 | AC005302 | AC005302 Mus muscu |
| 24 | c | 40.2 | 8.9 | 47852 | 1 | MTV023 | AL022022 Mycobacte |
| 25 | c | 40.2 | 8.9 | 213455 | 32 | AL135901 | AL135901 Homo sapi |
| 26 | c | 40.2 | 8.9 | 227070 | 56 | AC023888 | AC023888 Homo sapi |
| 27 | | 40 | 8.8 | 5880 | 16 | HSE4 | D14486 Equine herp |
| 28 | | 40 | 8.8 | 54775 | 55 | AC023526 | AC023526 Homo sapi |
| 29 | | 40 | 8.8 | 121882 | 44 | AC019147 | AC019147 Homo sapi |
| 30 | | 39.8 | 8.8 | 11320 | 3 | CVCMYC | X95367 C.familiari |
| 31 | | 39.8 | 8.8 | 10167 | 2 | SFU08223 | U08223 Streptomyce |
| 32 | c | 39.8 | 8.8 | 44300 | 54 | AC022749 | AC022749 Homo sapi |
| 33 | c | 39.8 | 8.8 | 116789 | 54 | AC015462 | AC015462 Homo sapi |
| 34 | | 39.8 | 8.8 | 227178 | 55 | AC023492 | AC023492 Homo sapi |
| 35 | | 39.6 | 8.7 | 121882 | 44 | AC019147 | AC019147 Homo sapi |
| 36 | c | 39.6 | 8.7 | 126650 | 44 | AC019049 | AC019049 Homo sapi |
| 37 | | 39.4 | 8.7 | 806 | 7 | HVU49482 | U49482 Hordeum vul |
| 38 | | 39.4 | 8.7 | 2578 | 12 | MUSHOXAA | L08757 Mus musculu |
| 39 | | 39.4 | 8.7 | 51950 | 55 | AC023561 | AC023561 Homo sapi |
| 40 | c | 39.4 | 8.7 | 127778 | 44 | AC020735 | AC020735 Homo sapi |
| 41 | | 39.2 | 8.7 | 2157 | 4 | XELB1AA | M63663 X.borealis |
| 42 | | 39.2 | 8.7 | 3252 | 10 | HSTAFIII13 | Y11354 H.sapiens m |
| 43 | c | 39.2 | 8.7 | 42189 | 11 | AC005258 | AC005258 Homo sapi |
| 44 | | 39.2 | 8.7 | 69798 | 54 | AC022682 | AC022682 Mus muscu |
| 45 | c | 39.2 | 8.7 | 102713 | 44 | AC019126 | AC019126 Homo sapi |

ALIGNMENTS

```
RESULT     1
AC008763
LOCUS       AC008763    214575 bp    DNA        HTG    26-JAN-2000
DEFINITION  Homo sapiens chromosome 19 clone CTD-3214H19, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC008763
VERSION     AC008763.2   GI:6759002
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 214575)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 214575)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            On Jan 26, 2000 this sequence version replaced gi:5686308.
COMMENT     ------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov

* NOTE: This record contains 143 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

*          1     526: contig of 526 bp in length
*        527    1172: gap of unknown length
*       1173    1874: contig of 702 bp in length
*       1875    2526: gap of unknown length
*       2527    3341: contig of 815 bp in length
*       3342    3980: gap of unknown length
*       3981    4621: contig of 641 bp in length
*       4622    5267: gap of unknown length
*       5268    5965: contig of 698 bp in length
*       5966    6499: gap of unknown length
*       6500    6802: contig of 303 bp in length
*       6803    7278: gap of unknown length
*       7279    7794: contig of 516 bp in length
*       7795    8099: gap of unknown length
*       8100    8412: contig of 313 bp in length
*       8413    8977: gap of unknown length
*       8978    9609: contig of 632 bp in length
*       9610   10398: contig of 789 bp in length
*      10399   11373: gap of unknown length
*      11374   12245: contig of 872 bp in length
*      12246   12994: gap of unknown length
*      12995   13817: contig of 823 bp in length
*      13818   14834: gap of unknown length
*      14835   15867: contig of 1033 bp in length
*      15868   16456: gap of unknown length
*      16457   17173: contig of 717 bp in length
*      17174   18365: gap of unknown length
*      18366   19087: contig of 722 bp in length
*      19088   19955: gap of unknown length
*      19956   20899: contig of 944 bp in length
*      20900   21726: gap of unknown length
*      21727   22451: contig of 725 bp in length
*      22452   23825: gap of unknown length
*      23826   24745: contig of 920 bp in length
*      24746   25345: gap of unknown length
*      25346   26371: contig of 1026 bp in length
*      26372   27436: gap of unknown length
*      27437   27957: contig of 521 bp in length
*      27958   28620: gap of unknown length
*      28621   29584: contig of 964 bp in length
*      29585   30443: gap of unknown length
*      30444   31147: contig of 704 bp in length
*      31148   32101: gap of unknown length
*      32102   32865: contig of 764 bp in length
*      32866   33541: gap of unknown length
*      33542   34142: contig of 601 bp in length
*      34143   35052: gap of unknown length
*      35053   35799: contig of 747 bp in length
*      35800   36625: gap of unknown length
*      36626   36835: contig of 210 bp in length
*      36836   37430: gap of unknown length
*      37431   38409: contig of 979 bp in length
*      38410   38581: gap of unknown length
*      38582   38656: contig of 75 bp in length
```

```
RESULT 2
AC016932
LOCUS       AC016932     90442 bp    DNA             HTG       19-FEB-2000
DEFINITION  Homo sapiens clone RP11-166C10, *** SEQUENCING IN PROGRESS ***, 53
            unordered pieces.
ACCESSION   AC016932
VERSION     AC016932.2  GI:7007858
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 90442)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 90442)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Feb 19, 2000 this sequence version replaced gi:6552811.
COMMENT     ----- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information
            Center project name: HMTH
            Center clone name: RP11-166C10
            ----- Summary Statistics
            Sequencing vector: M13; L08821
            Chemistry: Dye-terminator Big Dye: 100% of reads
            Assembly program: Phrap; version 0.980611
            Consensus quality: 41098 bases at least Q40
            Consensus quality: 53818 bases at least Q30
            Consensus quality: 60396 bases at least Q20
            Estimated insert size: 90442; agarose-fp estimation
            Estimated insert size: 85446; sum-of-contigs estimation
            Quality coverage: 0.7x in Q20 bases; agarose-fp estimation
            Quality coverage: 0.7x in Q20 bases; sum-of-contigs estimation
            ------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 53 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            ------------
            *          1     811: contig of 811 bp in length
            *        812     831: gap of unknown length
```

```
*                      gap of unknown length
*        38657  39642: contig of 986 bp in length
*                      gap of unknown length
*        39643  40250: contig of 608 bp in length
*                      gap of unknown length
*        40251  40987: contig of 737 bp in length
*                      gap of unknown length
*        40988  41412: contig of 425 bp in length
*                      gap of unknown length
*        41413  42546: contig of 1134 bp in length
*                      gap of unknown length
*        42547  43214: contig of 668 bp in length
*                      gap of unknown length
*        43215  43642: contig of 428 bp in length
*                      gap of unknown length
*        43643  44051: contig of 409 bp in length
*                      gap of unknown length
*        44052  44837: contig of 786 bp in length
*                      gap of unknown length
*        44838  45722: contig of 885 bp in length
*                      gap of unknown length
*        45723  46024: contig of 302 bp in length
*                      gap of unknown length
*        46025  46792: contig of 768 bp in length
*                      gap of unknown length
*        46793  47552: contig of 770 bp in length
*                      gap of unknown length
*        47553  49152: contig of 1590 bp in length
*                      gap of unknown length
*        49153  50130: contig of 978 bp in length
*                      gap of unknown length
*        50131  51208: contig of 1078 bp in length
*                      gap of unknown length
*        51209  52105: contig of 897 bp in length
*                      gap of unknown length
*        52106  53375: contig of 1270 bp in length
*                      gap of unknown length
*        53376  54894: contig of 1519 bp in length
*                      gap of unknown length
*        54895  55139: contig of 245 bp in length
*                      gap of unknown length
*        55140  56348: contig of 1209 bp in length
*                      gap of unknown length
*        56349  57711: contig of 1363 bp in length
*                      gap of unknown length
*        57712  58273: contig of 562 bp in length
*                      gap of unknown length
*        58274  59036: contig of 763 bp in length
*                      gap of unknown length
*        59037  60392: contig of 1356 bp in length
*                      gap of unknown length
*        60393  61079: contig of 687 bp in length
*                      gap of unknown length
*        61080  61936: contig of 857 bp in length
*                      gap of unknown length
*        61937  62600: contig of 664 bp in length
*                      gap of unknown length
```

```
Query Match 28.5%;  Score 129;  DB 45;  Length 214575;
Best Local Similarity 100.0%;  Pred. No. 6.7e-19;
Matches 129;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    36 GCGCCTGCAGGATGAAAGCTCTCGTCCTCCTCCCTGTCCCGTGGGCGTGTTGGTGT 95
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 82661 GCGCCTGCAGGATGAAAGCTCTCGTCCTCCTCCCTGTCCCGTGGGCGTGTTGGTGT 82720

Qy    96 CTAGCAAGACCCTGCTCCATGGAAGAAGCCATCAATGAGAGATCCAGGAGGTCGCCG 155
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 82721 CTAGCAAGACCCTGCTCCATGGAAGAAGCCATCAATGAGAGGATCCAGGAGGTCGCCG 82780

Qy   156 GCTCCCTAA 164
         |||||||||
Db 82781 GCTCCCTAA 82789
```

```
*    832    1876:  contig of 1045 bp in length
*   1877    1896:  gap of unknown length
*   1897    2882:  contig of 986 bp in length
*   2883    2902:  gap of unknown length
*   2903    3993:  contig of 1091 bp in length
*   3994    4013:  gap of unknown length
*   4014    4811:  contig of 798 bp in length
*   4812    4831:  gap of unknown length
*   4832    6227:  contig of 1396 bp in length
*   6228    6247:  gap of unknown length
*   6248    7283:  contig of 1036 bp in length
*   7284    7303:  gap of unknown length
*   7304    8543:  contig of 1240 bp in length
*   8544    8563:  gap of unknown length
*   8564   10252:  contig of 1689 bp in length
*  10253   10272:  gap of unknown length
*  10273   11441:  contig of 1169 bp in length
*  11442   11461:  gap of unknown length
*  11462   12487:  contig of 1026 bp in length
*  12488   12507:  gap of unknown length
*  12508   13621:  contig of 1114 bp in length
*  13622   13641:  gap of unknown length
*  13642   15080:  contig of 1439 bp in length
*  15081   15100:  gap of unknown length
*  15101   16137:  contig of 1037 bp in length
*  16138   16157:  gap of unknown length
*  16158   17466:  contig of 1309 bp in length
*  17467   17486:  gap of unknown length
*  17487   18999:  contig of 1513 bp in length
*  19000   19019:  gap of unknown length
*  19020   20438:  contig of 1419 bp in length
*  20439   20458:  gap of unknown length
*  20459   21658:  contig of 1200 bp in length
*  21659   21678:  gap of unknown length
*  21679   22207:  contig of 529 bp in length
*  22208   22227:  gap of unknown length
*  22228   23695:  contig of 1468 bp in length
*  23696   23715:  gap of unknown length
*  23716   24806:  contig of 1091 bp in length
*  24807   24826:  gap of unknown length
*  24827   26551:  contig of 1725 bp in length
*  26552   26571:  gap of unknown length
*  26572   27965:  contig of 1394 bp in length
*  27966   27985:  gap of unknown length
*  27986   29263:  contig of 1278 bp in length
*  29264   29283:  gap of unknown length
*  29284   30520:  contig of 1237 bp in length
*  30521   30540:  gap of unknown length
*  30541   32236:  contig of 1696 bp in length
*  32237   32256:  gap of unknown length
*  32257   33693:  contig of 1437 bp in length
*  33694   33713:  gap of unknown length
*  33714   35327:  contig of 1614 bp in length
*  35328   35347:  gap of unknown length
*  35348   36131:  contig of 784 bp in length
*  36132   36151:  gap of unknown length
*  36152   38604:  contig of 2453 bp in length
*  38605   38624:  gap of unknown length
*  38625   40511:  contig of 1887 bp in length
*  40512   40531:  gap of unknown length
*  40532   42146:  contig of 1615 bp in length
*  42147   42166:  gap of unknown length
*  42167   43749:  contig of 1583 bp in length
*  43750   43769:  gap of unknown length
*  43770   45010:  contig of 1241 bp in length
*  45011   45030:  gap of unknown length
*  45031   46545:  contig of 1515 bp in length
*  46546   46565:  gap of unknown length
*  46566   48528:  contig of 1963 bp in length
*  48529   48548:  gap of unknown length
*  48549   50672:  contig of 2124 bp in length
*  50673   50692:  gap of unknown length
*  50693   52614:  contig of 1922 bp in length

*  52615   52634:  gap of unknown length
*  52635   54090:  contig of 1456 bp in length
*  54091   54110:  gap of unknown length
*  54111   56624:  contig of 2514 bp in length
*  56625   56644:  gap of unknown length
*  56645   59368:  contig of 2724 bp in length
*  59369   59388:  gap of unknown length
*  59389   62256:  contig of 2868 bp in length
*  62257   62276:  gap of unknown length
*  62277   64454:  contig of 2378 bp in length
*  64455   64474:  gap of unknown length
*  64555   64674:  contig of 2378 bp in length
*  64675   67214:  contig of 2540 bp in length
*  67215   67234:  gap of unknown length
*  67235   68685:  contig of 1451 bp in length
*  68686   68705:  gap of unknown length
*  68706   70177:  contig of 1473 bp in length
*  70178   70198:  gap of unknown length
*  70199   72544:  contig of 2346 bp in length
*  72545   72564:  gap of unknown length
*  72565   75021:  contig of 2457 bp in length
*  75022   75041:  gap of unknown length
*  75042   78251:  contig of 3210 bp in length
*  78252   78271:  gap of unknown length
*  78272   80614:  contig of 2343 bp in length
*  80615   80634:  gap of unknown length
*  80635   83941:  contig of 3307 bp in length
*  83942   83961:  gap of unknown length
*  83962   87800:  contig of 3839 bp in length
*  87801   87820:  gap of unknown length
*  87821   90442:  contig of 2622 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1..90442
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-166C10"
BASE COUNT    27023 a    18419 c    17550 g    26309 t    1141 others
ORIGIN

Query Match              15.4%;   Score 69.6;   DB 57;  Length 90442;
Best Local Similarity    72.6%;   Pred. No. 5.2e-06;
Matches   90;  Conservative   0;  Mismatches  34;  Indels   0;  Gaps   0;

Qy    241 AGGCTTCGCGTCACCGGCTCGTTGGCTCCGCCTGCTCGTGGCTCCGTGGGATGTGCGCGC 300
          ||| |-||-||||| |||| |||||||| - ||||||||| |||||||||||||||||| |
Db  20503 AGGGATGGCTGTCACTGGCTGTGCTTGGCTATGCGTGTTGTGCTGTGTTCGTGGGATGTTCAGCT 20562

Qy    301 CGAGACCACATGTCACTGCCAGTGCGCGGGCATGGACTGACGTGGCGGCGCTGCTGTCG 360
          || |||||| |-|||||||||||||||| | ||||||| |||||||| || |||||| |
Db  20563 GGAAACCACCTGCCACTGCCAGTGCCAGTGTGGTGGTGGACTGGACCACTGGACCACTGCCA 20622

Qy    361 TGTG 364
          ||||
Db  20623 CCTG 20626

RESULT  3
AC016932/c
LOCUS       AC016932    90442 bp    DNA            HTG       19-FEB-2000
DEFINITION  Homo sapiens clone RP11-166C10, *** SEQUENCING IN PROGRESS ***, 53
            unordered pieces.
ACCESSION   AC016932
VERSION     AC016932.2   GI:7007858
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 90442)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,Z., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
```

```
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,J., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 90442)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Feb 19, 2000 this sequence version replaced gi:6552811.
            -------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -------- Project Information
            Center project name: HMTH
            Center clone name: RP11-166C10
            -------- Summary Statistics
            Sequencing vector: M13; L08821
            Chemistry: Dye-terminator Big Dye: 100% of reads
            Assembly program: Phrap; version 0.980611
            Consensus quality: 41098 bases at least Q40
            Consensus quality: 53818 bases at least Q30
            Consensus quality: 60396 bases at least Q20
            Estimated insert size: 90442; agarose-fp estimation
            Estimated insert size: 85446; sum-of-contigs estimation
            Quality coverage: 0.7x in Q20 bases; agarose-fp estimation
            Quality coverage: 0.7x in Q20 bases; sum-of-contigs estimation
            --------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 53 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          * --------
          *     1    811: contig of 811 bp in length
          *   812    831: gap of unknown length
          *   832   1876: contig of 1045 bp in length
          *  1877   1896: gap of unknown length
          *  1897   2882: contig of 986 bp in length
          *  2883   2902: gap of unknown length
          *  2903   3993: contig of 1091 bp in length
          *  3994   4013: gap of unknown length
          *  4014   4811: contig of 798 bp in length
          *  4812   4831: gap of unknown length
          *  4832   6227: contig of 1396 bp in length
          *  6228   6247: gap of unknown length
          *  6248   7283: contig of 1036 bp in length
          *  7284   7303: gap of unknown length
          *  7304   8543: contig of 1240 bp in length
          *  8544   8563: gap of unknown length
          *  8564  10252: contig of 1689 bp in length
          * 10253  10272: gap of unknown length
          * 10273  11441: contig of 1169 bp in length
          * 11442  11461: gap of unknown length
          * 11462  12487: contig of 1026 bp in length

          * 12488  12507: gap of unknown length
          * 12508  13621: contig of 1114 bp in length
          * 13622  13641: gap of unknown length
          * 13642  15080: contig of 1439 bp in length
          * 15081  15100: gap of unknown length
          * 15101  16137: contig of 1037 bp in length
          * 16138  16157: gap of unknown length
          * 16158  17466: contig of 1309 bp in length
          * 17467  17486: gap of unknown length
          * 17487  18999: contig of 1513 bp in length
          * 19000  19019: gap of unknown length
          * 19020  20438: contig of 1419 bp in length
          * 20439  20458: gap of unknown length
          * 20459  21658: contig of 1200 bp in length
          * 21659  21678: gap of unknown length
          * 21679  22207: contig of 529 bp in length
          * 22208  22227: gap of unknown length
          * 22228  23695: contig of 1468 bp in length
          * 23696  23715: gap of unknown length
          * 23716  24806: contig of 1091 bp in length
          * 24807  24826: gap of unknown length
          * 24827  26551: contig of 1725 bp in length
          * 26552  26571: gap of unknown length
          * 26572  27965: contig of 1394 bp in length
          * 27966  27985: gap of unknown length
          * 27986  29263: contig of 1278 bp in length
          * 29264  29283: gap of unknown length
          * 29284  30520: contig of 1237 bp in length
          * 30521  30540: gap of unknown length
          * 30541  32236: contig of 1696 bp in length
          * 32237  32256: gap of unknown length
          * 32257  33693: contig of 1437 bp in length
          * 33694  33713: gap of unknown length
          * 33714  35327: contig of 1614 bp in length
          * 35328  35347: gap of unknown length
          * 35348  36131: contig of 784 bp in length
          * 36132  36151: gap of unknown length
          * 36152  38604: contig of 2453 bp in length
          * 38605  38624: gap of unknown length
          * 38625  40511: contig of 1887 bp in length
          * 40512  40531: gap of unknown length
          * 40532  42146: contig of 1615 bp in length
          * 42147  42166: gap of unknown length
          * 42167  43749: contig of 1583 bp in length
          * 43750  43769: gap of unknown length
          * 43770  45010: contig of 1241 bp in length
          * 45011  45030: gap of unknown length
          * 45031  46545: contig of 1515 bp in length
          * 46546  46565: gap of unknown length
          * 46566  48528: contig of 1963 bp in length
          * 48529  48548: gap of unknown length
          * 48549  50672: contig of 2124 bp in length
          * 50673  50692: gap of unknown length
          * 50693  52614: contig of 1922 bp in length
          * 52615  52634: gap of unknown length
          * 52635  54090: contig of 1456 bp in length
          * 54091  54110: gap of unknown length
          * 54111  56624: contig of 2514 bp in length
          * 56625  56644: gap of unknown length
          * 56645  59368: contig of 2724 bp in length
          * 59369  59388: gap of unknown length
          * 59389  62256: contig of 2868 bp in length
          * 62257  62276: gap of unknown length
          * 62277  64654: contig of 2378 bp in length
          * 64655  64674: gap of unknown length
          * 64675  67214: contig of 2540 bp in length
          * 67215  67234: gap of unknown length
          * 67235  68685: contig of 1451 bp in length
          * 68686  68705: gap of unknown length
          * 68706  70178: contig of 1473 bp in length
          * 70179  70198: gap of unknown length
          * 70199  72544: contig of 2346 bp in length
          * 72545  72564: gap of unknown length
```

```
                                                    SVSTSHTSMSGTSSRGGRYGSGGGGGTYGGSRGSYGGSSSGGGS
                                                    GGGSYGGSGSSGHRGSGGGSSGGGSSYGGSGSSSGGRGSSSSGGGGVKSGSSSSVKFVST
                                                    TYSRGTN"
                            polyA_signal     217 a      958..963
                            BASE COUNT               265 c      289 g       210 t
                            ORIGIN

                            Query Match                  9.7%;     Score 43.8;  DB 12;  Length 981;
                            Best Local Similarity       45.0%;     Pred. No. 3.5;
                            Matches 165; Conservative    0;  Mismatches  202;  Indels    0;  Gaps    0;

                        Qy    69 TCCTCCCTGTCCTGGGCTGCTGTTGGTGTCTTAGCAAGACCCTGTCCTCCATGGAAGAAGCCA 128
                                  | |||| ||  ||  |||| |||| || |  |||||| || | |||| |||||||||||||
                        Db     5 TGCTGCGTGACTACCAGGAGCTGATGAACACCAAGCTGCCCTGCCCCTGACATGGAGATCGCCA  64

                        Qy   129 TCAATGAGAGGATCCAGGAGGTGCCGCGGCTCCTAATATTAGGGCAATAAGCAGCATTG 188
                                  |||| | || || |||| || |||| |||| || ||| || ||| ||||| || ||
                        Db    65 CCTACAGGAAACTTCTGAAGGAGGAGGAGATCAGGATGTCTGGAGAATGCACCCCCAACG 124

                        Qy   189 GCCTGGGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTACTTGCCCCGAGGGCTTCG 248
                                  ||  || |  || ||  || |||| | |||| |||| || ||  |||||| ||||||| |
                        Db   125 TGAGCGTGTCTGTGAGCACCAGCCACACTAGCATGCATGAGCGGAACGAGCAGCGAGGTGGCG 184

                        Qy   249 CCGTCACCGGCTGCACTTGTGGCTCCGCCTGTGCCTGCGTGGGGATGTGCGCGCCGAGACCA 308
                                  || | || ||  |||| ||| || |||| | ||| || |||||| || ||| || |||
                        Db   185 GCAGATACGGCTCCGGAGCGGCGGCGGTGCCGGGCGCGGGGCTCGTCGTGTGCAGC 244

                        Qy   309 CATGTCACTGCCAGTGCCGGCGGCATGGACCGGAGCGGCGGCT 368
                                  || ||  |  ||  ||| |||| || || || |||||| | |||
                        Db   245 GAGCTACGGAGCGGGTTCCGAGGCGGCAGCTACGGAGGTTGCAGCCTTCGGAGCCTCGTGTGCAGC 304

                        Qy   369 CCTGAGGTCGCGCGGCGGCAGCGCCGTGCACAGCGGCGGGCGGAGCGGGCTCCAGTCCGGAGGGG 428
                                  || ||| |  ||  ||   ||||| ||||||||| |||| ||||  |||||||| |||||||||
                        Db   305 CCGGAGGCGGCGAGCTACGGCGGCTACGGCGGCCACCAGAGGTGGCGGCCACAGAGGTGCTCTGGAGGGG 364

                        Qy   429 TTGCGGG 435
                                  |||||||
                        Db   365 GCGGTGG 371

                        RESULT  5
                        SC6G4
                        LOCUS         SC6G4       41055 bp    DNA                          BCT       20-AUG-1998
                        DEFINITION    Streptomyces coelicolor cosmid 6G4.
                        ACCESSION     AL031317
                        VERSION       AL031317.1    GI:3449234
                        KEYWORDS      10 kD chaperonin; 30S ribosomal protein S11; 30S ribosomal protein
                                      S13; 30S ribosomal protein S9; 50S ribosomal protein L13; 50S
                                      ribosomal protein L17; 50S ribosomal protein L36; 60 kD chaperonin;
                                      acetyltransferase; acpS; adenylate kinase; adk; alanine racemase;
                                      alr; coaA; cpn10; cpn60; dehydrogenase; DNA-directed RNA polymerase
                                      alpha chain; gcp; glmS; glucosamine--fructose-6-phosphate
                                      aminotransferase; groEL1; groES; holo-[acyl-carrier protein]
                                      synthase; infA; lipase; map; methionine aminopeptidase;
                                      O-sialoglycoprotein endopeptidase; pantothenate kinase;
                                      phospho-sugar mutase; pseudouridylate synthase; rplM; rplQ; rpmJ;
                                      rpoA; rpsI; rpsK; rpsM; translational initiation factor IF1; truA;
                                      two-component regulator.
                        SOURCE        Streptomyces coelicolor A3(2).
                          ORGANISM    Streptomyces coelicolor A3(2)
                                      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
                                      Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
                        REFERENCE     1 (bases 1 to 41055)
                          AUTHORS     Saunders,D. and Harris,D.
                          JOURNAL     Unpublished
                        REFERENCE     2 (bases 1 to 41055)
                          AUTHORS     Parkhill,J., Barrell,B.G. and Rajandream,M.A.
                          TITLE       Direct Submission
                          JOURNAL     Submitted (19-AUG-1998) Streptomyces coelicolor sequencing project,
                                      Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
                                      CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
                                      David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
```

```
                        *     72565      75021: contig of 2457 bp in length
                        *     75022      75041: gap of unknown length
                        *     75042      78251: contig of 3210 bp in length
                        *     78252      78271: gap of unknown length
                        *     78272      80614: contig of 2343 bp in length
                        *     80615      80634: gap of unknown length
                        *     80635      83941: contig of 3307 bp in length
                        *     83942      83961: gap of unknown length
                        *     83962      87800: contig of 3839 bp in length
                        *     87801      87820: gap of unknown length
                        *     87821      90442: contig of 2622 bp in length.
                        FEATURES             Location/Qualifiers
                             source          1..90442
                                             /organism="Homo sapiens"
                                             /db_xref="taxon:9606"
                                             /clone="RP11-166C10"

                        BASE COUNT   27023 a   18419 c   17550 g   26309 t   1141 others
                        ORIGIN

                        Query Match                 10.1%;    Score 45.8;  DB 57;  Length 90442;
                        Best Local Similarity       62.8%;    Pred. No. 0.74;
                        Matches 71; Conservative     0;  Mismatches  42;  Indels    0;  Gaps    0;

                        Qy   241 AGGCTTCGCCGTCACCGGCTGCACTTGTGGCTCCGCCTGTGGCTCGTCGTGGATGTGCGCGC 300
                                  |||  ||| |||||||| |||||| | | ||||| || |||||||||||||||
                        .Db   262 AGGGTACCCTGTCACTGGCTCATGTCATGCCCGGAAAAGGCTGTGGCTCCTGGGATGTCCTATG 203

                        Qy   301 CGAGACCACATGTCACTGCCAGTGCCGCGGGCATGGACCGGAGCGGCGGCT 353
                                  || ||| | ||||||||||||||||| ||||||||| ||| |||| || |
                        Db   202 GGAATCCATCTGTCACTGCCAGTGTGTTGTTTTGGATGGGTCACTGCCTGCT 150

                        RESULT  4
                        RNCKERIMP
                        LOCUS         RNCKERIMP    981 bp    mRNA            ROD       30-APR-1992
                        DEFINITION    R.norvegicus mRNA for cytokeratin type I (3' end).
                        ACCESSION     X54806
                        VERSION       X54806.1    GI:55955
                        KEYWORDS      cytokeratin; cytokeratin type I.
                        SOURCE        Norway rat.
                          ORGANISM    Rattus norvegicus
                                      Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
                                      Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
                        REFERENCE     1 (bases 1 to 981)
                          AUTHORS     Redfern,C.P.F.
                          TITLE       Direct Submission
                          JOURNAL     Submitted (26-SEP-1990) Redfern C.P.F., Univ. of Newcastle upon
                                      Tyne, Medical Mol. Biology Group, Dept. of Dermathology, 4th Floor
                                      Cookson Bldg, Medical School, Framlington Place, Newcastle upon
                                      Tyne NE2 4HH, UK
                        REFERENCE     2 (bases 1 to 981)
                          AUTHORS     Redfern,C.P.F. and Allen,G.
                          TITLE       Sequence variation (between species) at the carboxyterminal domain
                                      of cytokeratin-1
                          JOURNAL     Unpublished
                        FEATURES             Location/Qualifiers
                             source          1..981
                                             /organism="Rattus norvegicus"
                                             /strain="Wistar"
                                             /db_xref="taxon:10116"
                                             /dev_stage="1-2 day neonates"
                                             /tissue_type="epidermis"
                                             /cell_type="keratinocytes"
                                             /clone_lib="plasmid, epidermal cDNA"
                                             /clone="p.EL3.25"
                             CDS             <1..504
                                             /codon_start=1
                                             /product="cytokeratin type I"
                                             /protein_id="CAA38577.1"
                                             /db_xref="GI:55956"
                                             /db_xref="SPTREMBL:Q63115"
                                             /translation="RLLRDYQELMNTKLALDMEIATYRKLLEGEIRMSGECTPNVSV
```

REFERENCE   3 (bases 1 to 41055)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE   A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
COMMENT   Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC.
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
            numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon. Gene
            prediction is based on positional base preference in codons using a
            specially developed Hidden Markov Model (Krogh et al., Nucleic
            Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
            of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
            correct initiation codon. Where possible we choose an initiation
            codon (atg, gtg, ttg or gtt)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon). If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone. It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions. Cosmid 6G4 lies
            between D31 and D63 in the AseI-D genomic restriction fragment.
FEATURES   Location/Qualifiers
     source   1..41055
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
                     /clone="cosmid 6G4"
     misc_feature   1..1610
                     /note="98.4% match to SCSECYDNA X83011 S.coelicolor secY
                     locus DNA from 4547 to 6154"
     gene   1..286
                     /gene="adk"
     CDS   <1..286
                     /gene="adk"
                     /note="SC6G4.01, adk, adenylate kinase, partial CDS, len:
                     >94 aa; almost identical to KAD_STRCO adenylate kinase (EC
                     2.7.4.3) (205 aa) and highly similar to many e.g.
                     KAD_BORPE adenylate kinase (EC 2.7.4.3)(218 aa), fasta
                     scores; opt: 253 z-score: 429.1 E(): 1.2e-16, 50.6%
                     identity in 85 aa overlap. Contains Pfam match to entry
                     PF00406 adenylatekinase, Adenylate kinases, score 136.80,
                     E-value 3.2e-37"
                     /codon_start=2
                     /transl_table=11
                     /product="adenylate kinase"
                     /protein_id="CAA20379.1"
                     /db_xref="GI:3449235"
                     /db_xref="SPTREMBL:O86771"
                     /translation="IAGRRVCRNEPKHVFHVTYTPPKEGVCDVCGGELYQRDDSEE
                     TVRRLEVYHTQTEPIIDYYKSQGLVATIAATGPVDEVTRRALEALKRDQ"
     misc_feature   2..208
                     /gene="adk"
                     /note="Pfam match to entry PF00406 adenylatekinase,
                     Adenylate kinases, score 136.80, E-value 3.2e-37"
     RBS   423..426
                     /note="possible RBS upstream of map"
     gene   438..1274
                     /gene="map"
     CDS   438..1274
                     /gene="map"
                     /note="SC6G4.02, map, methionine aminopeptidase, len: 278
                     aa; identical to TR:O54208 (EMBL:X83011) methionine
                     aminopeptidase (278 aa) and highly similar to many
                     e.g.AMPM_BACSU methionine aminopeptidase (EC 3.4.11.18)
                     (248 aa), fasta scores; opt: 546 z-score: 922.2 E(): 0,
                     44.5% identity in 256 aa overlap. Also similar to S.
                     coelicolor map2 (E(): 3.2e-33, 46.8% identity in 267 aa
                     overlap). Contains PS00680 Methionine aminopeptidase
                     subfamily 1 signature and Pfam match to entry PF00557
                     pep_M24, metallopeptidase family M24, score 194.80,
                     E-value 1.4e-54"
                     /codon_start=1
                     /transl_table=11
                     /product="methionine aminopeptidase"
                     /protein_id="CAA20380.1"
                     /db_xref="GI:3449236"
                     /db_xref="SPTREMBL:O54208"
                     /translation="MVOIKNPEQIAKMREAGLVVAAIHAATREAAVPGATTKDLDQVA
                     RKVLAEHDAKPNFLGVGGFPAICTSVNEVVHGIPSDDVLKDGDVISIDCGAIIDG
                     WHGDAATAFVGSGHSPELVELSRVTESMWAGIAAMKQGNRLVDVSRAIETYIRRQP
                     KPGGGKYGIIEDYGGHGIGTEMHMDPHLLNYVDRRRGKGPKLVPGFCLAIEPMVSLGT
                     PRTEVLPDEWTVITTDGTWSSHWEHSVALTEQGPLVLTSPDDGGKAKLAELGITRAAPDP
                     LA"
     misc_feature   444..1196
                     /gene="infA"
                     /note="Pfam match to entry PF00557 pep_M24,
                     metallopeptidase family M24, score 194.80, E-value
                     1.4e-54"
     misc_feature   954..1010
                     /gene="map"
                     /note="PS00680 Methionine aminopeptidase subfamily 1
                     signature"
     RBS   1405..1408
                     /gene="infA"
                     /note="possible RBS upstream of infA"
     gene   1416..1637
                     /gene="infA"
     CDS   1416..1637
                     /gene="infA"
                     /note="SC6G4.03, infA, translational initiation factor
                     IF1, len: 73 aa; identical to TR:O54209 (EMBL:X83011)
                     translational initiation factor IF1 (fragment) (65 aa) and
                     highly similar to many e.g. IF1_BACSU translation
                     initiation factor IF-1 (71 aa), fasta scores; opt: 384
                     z-score: 792.5 E(): 0, 78.6% identity in 70 aa overlap.
                     Contains Pfam match to entry PF00575 S1, S1 RNA binding
                     motif, score 54.10, E-value 3.1e-12"
                     /codon_start=1
                     /transl_table=11
                     /product="translational initiation factor IF1"
                     /protein_id="CAA20381.1"
                     /db_xref="GI:3449237"
                     /translation="MAKKQGAIEIEGTVVESLPNAMFKVELQNGHQVLAHISGKMRMH
                     YIRILPDDRVVVELSPYDLTRGRIVYRYK"
     misc_feature   1422..1634
                     /gene="infA"
                     /note="Pfam match to entry PF00575 S1, S1 RNA binding
                     motif, score 54.10, E-value 3.1e-12"
     RBS   1690..1693
                     /note="possible RBS upstream of rpmJ"
     misc_feature   1706..1816
                     /gene="rpmJ"
                     /note="Pfam match to entry PF00444 L36, Ribosomal protein
                     L36, score 70.40, E-value 3.1e-22"
     gene   1706..1819
                     /gene="rpmJ"
     CDS   1706..1819
                     /gene="rpmJ"

```
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1    1230: contig of 1230 bp in length
*     1231    1248: gap of unknown length
*     1249    2396: contig of 1148 bp in length
*     2397    2414: gap of unknown length
*     2415    3787: contig of 1373 bp in length
*     3788    3805: gap of unknown length
*     3806    4800: contig of 995 bp in length
*     4801    4818: gap of unknown length
*     4819    6003: contig of 1185 bp in length
*     6004    6021: gap of unknown length
*     6022    7046: contig of 1025 bp in length
*     7047    7064: gap of unknown length
*     7065    8363: contig of 1299 bp in length
*     8364    8381: gap of unknown length
*     8382    9757: contig of 1376 bp in length
*     9758    9775: gap of unknown length
*     9776   11228: contig of 1453 bp in length
*    11229   11246: gap of unknown length
*    11247   12288: contig of 1042 bp in length
*    12289   12306: gap of unknown length
*    12307   13335: contig of 1029 bp in length
*    13336   13353: gap of unknown length
*    13354   14464: contig of 1111 bp in length
*    14465   14482: gap of unknown length
*    14483   16058: contig of 1576 bp in length
*    16059   16076: gap of unknown length
*    16077   17212: contig of 1136 bp in length
*    17213   17230: gap of unknown length
*    17231   18544: contig of 1314 bp in length
*    18545   18562: gap of unknown length
*    18563   26836: contig of 8274 bp in length
*    26837   26854: gap of unknown length
*    26855   72428: contig of 45574 bp in length
*    72429   72445: gap of unknown length
*    72446   202697: contig of 130252 bp in length
*   202698  202714: gap of unknown length
*   202715  203972: contig of 1258 bp in length
*   203973  203989: gap of unknown length
*   203990  205220: contig of 1231 bp in length
*   205221  205237: gap of unknown length
*   205238  206308: contig of 1071 bp in length
*   206309  206325: gap of unknown length
*   206326  207651: contig of 1326 bp in length.

FEATURES             Location/Qualifiers
     source          1..207651
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-226I21"

BASE COUNT    60555 a   41188 c   43884 g   57785 t   4239 others
ORIGIN

Query Match            9.4%;   Score 42.8;  DB 55;  Length 207651;
Best Local Similarity 48.2%;   Pred. No. 3;
Matches  110;  Conservative   0;  Mismatches 118;  Indels  0;  Gaps  0;

Qy   216 GGGGGGACCTGGCTACTTGCCCGTAGGCTTCGCCCGAGGCTTGCCGGCTGCACTTGTGGCTCCG 275
         |||||| |||||| | ||||||| |||||  ||||| |||  |||||| ||| |||||| |||||
Db 204396 GGGGGGGGGGGGGGGGGGGGCCCCCCCCCCCCCCAGCCTCGTTTCGAAATGCCGGNGGCGGGCG 204455

Qy   276 CCTGTGGCTCGTCGGATGTGCGCGCCGGAGACCACATGTCACTGCGGGGGGCGGGGCATGG 335
         | | ||||| | | ||||| | ||| | |||| | ||| | |||||| |||||| |||||||
Db 204456 GCGCGGGGCCNCGCGCGGGGGGNGGGGGGCGGCGCGCCGGCGGGGGGGGCGGGGCGCGGCCG 204515

Qy   336 ACTGGACCGGACGCGGCGCTGCTGTCGTGCAGCCTGAGGTCGCCGGGCAGCGGCGTGCACA 395
         ||||| |||| ||||||| |||  | ||  |||||||| ||| ||| |||||| ||||||||
Db 204516 CCGCGGCGGGGGGGGGGGGGGGCCCCCCCCCAGCCTCCGGGGGGGGGGCGCCCGCGCCCGNCC 204575
```

```
Qy     396  GCGCGGGCGGAGGCGGCTCCAGGTCCGGAGGGGTTGCGGGAGCTGG 443
            || |||||| || | || || | || || || | || || ||
Db  204576  GCCGGGGCGGCGGCGGCGGGCGGNCGGCGCCCGGCGGCCGCGGGGCGCGGGGG 204623
```

RESULT 7
AC009974/c

```
LOCUS       AC009974   219565 bp   DNA              HTG       01-FEB-2000
DEFINITION  Homo sapiens clone RP11-459I19, WORKING DRAFT SEQUENCE, 26
            unordered pieces.
ACCESSION   AC009974
VERSION     AC009974.4  GI:6850611
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 219565)
AUTHORS     Waterston,R.H.
TITLE       The sequence of Homo sapiens clone
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 219565)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (08-SEP-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Feb 1, 2000 this sequence version replaced gi:6139211.

            --------------  Genome Center  --------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            --------- Project Information ---------
            Center project name: H_NH0459I19
            --------- Summary Statistics ---------
            Sequencing vector: M13; 95%
            Sequencing vector: plasmid; 5%
            Chemistry: Dye-primer ET; 76% of reads
            Chemistry: Dye-terminator Big Dye; 24% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 187993 bases at least Q40
            Consensus quality: 192413 bases at least Q30
            Consensus quality: 194956 bases at least Q20
            Insert size: 221000; agarose-fp
            Insert size: 219565; sum-of-contigs
            Quality coverage: 4.02 in Q20 bases; agarose-fp
            Quality coverage: 4.05 in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1     1431: contig of 1431 bp in length
            *      1432     2603: gap of unknown length
            *      2604     3741: contig of 1172 bp in length
            *      3742     4974: gap of unknown length
            *      4975     6393: contig of 1419 bp in length
            *      6394     7811: gap of unknown length
            *      7812     9125: contig of 1418 bp in length
            *      9126    10219: gap of unknown length
            *     10220    11675: contig of 1314 bp in length
            *     11676    12860: gap of unknown length
            *     12861    13926: contig of 1094 bp in length
            *     13927    15365: gap of unknown length
            *     15366    16580: contig of 1456 bp in length
            *     16581    17895: gap of unknown length
            *     17896    19237: contig of 1185 bp in length
            *     19238    20608: gap of unknown length
            *     20609    22858: contig of 1066 bp in length
            *     22859    25391: gap of unknown length
            *     25392    33525: contig of 1439 bp in length
            *     33526    45483: gap of unknown length
            *     45484    58111: contig of 1215 bp in length
            *     58112    76537: gap of unknown length
            *     76538   110392: contig of 1315 bp in length
            *    110393   142425: gap of unknown length
            *    142426   173482: contig of 1342 bp in length
            *    173483   219565: gap of unknown length
                                  contig of 1371 bp in length
                                  gap of unknown length
                                  contig of 2250 bp in length
                                  gap of unknown length
                                  contig of 2533 bp in length
                                  gap of unknown length
                                  contig of 8134 bp in length
                                  gap of unknown length
                                  contig of 11958 bp in length
                                  gap of unknown length
                                  contig of 12628 bp in length
                                  gap of unknown length
                                  contig of 18426 bp in length
                                  gap of unknown length
                                  contig of 33855 bp in length
                                  gap of unknown length
                                  contig of 32033 bp in length
                                  gap of unknown length
                                  contig of 31057 bp in length
                                  gap of unknown length
                                  contig of 46083 bp in length

FEATURES             Location/Qualifiers
     source          1..219565
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-459I19"

BASE COUNT    56041 a   52428 c   53875 g   53924 t   3297 others
ORIGIN

Query Match            9.4%;
Best Local Similarity  54.2%;  Score 42.8; DB 53; Length 219565;
Matches  83;  Conservative  0;  Mismatches  70;  Indels  0;  Gaps  0;

Qy    281  GGCTCGTGGGATGTGCGCGCCGAGACCACATGCCAGTGCGCGGGCATGGACTGG 340
           |||||||| ||||||  ||||| ||| | ||||| |||||||| || ||||  |||
Db  11568  GGCGCGGCGCGTCTGGGCGCCGTCNTCGGCGCCGCCGCCGCGGGCGTCGGCGGGGGTC 11509

Qy    341  ACCGGAGGCGGCTGCTGTCGTGTGCAGCCCTGCGGCGGCGCGTGCACAGCGCG 400
           ||||||  || || |||| | ||||  |||| || ||| ||||| ||| |  ||
Db  11508  GCGTCGGCGGCTCCGNGGGGGCGGGGCGCCGCGGGGCGGCGCGGGGCCCTCCGGC 11449

Qy    401  GGCGGAGGCGGCTCCAGGTCCGGAGGGGTTGCG 433
           |||||  ||| |||| || ||  ||||  || |||
Db  11448  GGCGGCGCGGCNGCCGCCGGTCGGCGGCTGGGGGCGGCG 11416

RESULT 8
AC022648
LOCUS       AC022648   76968 bp   DNA              HTG       06-FEB-2000
DEFINITION  Homo sapiens clone RP11-28G14, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC022648
VERSION     AC022648.1  GI:6910660
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 76968)
```

AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE   Homo sapiens, clone RP11-28G14
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 76968)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
           Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
           Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
           Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
           DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
           Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
           Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
           Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
           Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
           Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
           McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
           Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
           Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
           Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
           Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
           Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
           Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           ---------- Genome Center
             Center: Whitehead Institute/ MIT Center for Genome Research
             Center code: WIBR
             Web site: http://www-seq.wi.mit.edu
             Contact: sequence_submissions@genome.wi.mit.edu
           ---------- Project Information
             Center project name: L4753
             Center clone name: 28_G_14
           ----
           * NOTE: This record contains 84 individual
           * sequencing reads that have not been assembled into
           * contigs. Runs of N are used to separate the reads
           * and the order in which they appear is completely
           * arbitrary. Low-pass sequence sampling is useful for
           * identifying clones that may be gene-rich and allows
           * overlap relationships among clones to be deduced.
           * However, it should not be assumed that this clone
           * will be sequenced to completion. In the event that
           * the record is updated, the accession number will
           * be preserved.
           *   1   934: contig of 934 bp in length
           *   gap of unknown length
           *   935   1846: contig of 912 bp in length
           *   gap of unknown length
           *   1847   2767: contig of 921 bp in length
           *   gap of unknown length
           *   2768   3689: contig of 922 bp in length
           *   gap of unknown length
           *   3690   4613: contig of 924 bp in length
           *   gap of unknown length
           *   4614   5528: contig of 915 bp in length
           *   gap of unknown length
           *   5529   6442: contig of 914 bp in length
           *   gap of unknown length
           *   6443   7262: contig of 820 bp in length
           *   gap of unknown length
           *   7263   8207: contig of 945 bp in length
           *   gap of unknown length
           *   8208   9135: contig of 928 bp in length
           *   gap of unknown length
           *   9136   10033: contig of 898 bp in length
           *   gap of unknown length
           *   10034   10959: contig of 926 bp in length
           *   gap of unknown length
           *   10960   11874: contig of 915 bp in length
           *   gap of unknown length
           *   11875   12780: contig of 906 bp in length
           *   gap of unknown length
           *   12781   13673: contig of 893 bp in length
           *   gap of unknown length
           *   13674   14593: contig of 920 bp in length
           *   gap of unknown length
           *   14594   15516: contig of 923 bp in length
           *   gap of unknown length
           *   15517   16460: contig of 944 bp in length
           *   gap of unknown length
           *   16461   17389: contig of 929 bp in length
           *   gap of unknown length
           *   17390   18312: contig of 923 bp in length
           *   gap of unknown length
           *   18313   19216: contig of 904 bp in length
           *   gap of unknown length
           *   19217   20158: contig of 942 bp in length
           *   gap of unknown length
           *   20159   21083: contig of 925 bp in length
           *   gap of unknown length
           *   21084   21979: contig of 896 bp in length
           *   gap of unknown length
           *   21980   22886: contig of 907 bp in length
           *   gap of unknown length
           *   22887   23781: contig of 895 bp in length
           *   gap of unknown length
           *   23782   24642: contig of 861 bp in length
           *   gap of unknown length
           *   24643   25576: contig of 934 bp in length
           *   gap of unknown length
           *   25577   26512: contig of 936 bp in length
           *   gap of unknown length
           *   26513   27426: contig of 914 bp in length
           *   gap of unknown length
           *   27427   28310: contig of 884 bp in length
           *   gap of unknown length
           *   28311   29266: contig of 956 bp in length
           *   gap of unknown length
           *   29267   30196: contig of 930 bp in length
           *   gap of unknown length
           *   30197   31100: contig of 904 bp in length
           *   gap of unknown length
           *   31101   32008: contig of 908 bp in length
           *   gap of unknown length
           *   32009   32885: contig of 877 bp in length
           *   gap of unknown length
           *   32886   33775: contig of 890 bp in length
           *   gap of unknown length
           *   33776   34653: contig of 878 bp in length
           *   gap of unknown length
           *   34654   35589: contig of 936 bp in length
           *   gap of unknown length
           *   35590   36500: contig of 911 bp in length
           *   gap of unknown length
           *   36501   37414: contig of 914 bp in length
           *   gap of unknown length
           *   37415   38364: contig of 950 bp in length
           *   gap of unknown length
           *   38365   39293: contig of 929 bp in length
           *   gap of unknown length
           *   39294   40220: contig of 927 bp in length
           *   gap of unknown length
           *   40221   41133: contig of 913 bp in length
           *   gap of unknown length
           *   41134   42043: contig of 910 bp in length
           *   gap of unknown length
           *   42044   42954: contig of 911 bp in length
           *   gap of unknown length
           *   42955   43867: contig of 913 bp in length
           *   gap of unknown length
           *   43868   44790: contig of 923 bp in length
           *   gap of unknown length
           *   44791   45714: contig of 924 bp in length

SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 164520)
AUTHORS   Waterston,R.H.
TITLE   The sequence of Homo sapiens clone
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 164520)
AUTHORS   Waterston,R.H.
TITLE   Direct Submission
JOURNAL   Submitted (08-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT   Center project name: H_NR0753I14.
            * NOTE: This record contains 231 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *   1   707: contig of 706 bp in length
            *   707   716: gap of unknown length
            *   717   1234: contig of 518 bp in length
            *   1235   1244: gap of unknown length
            *   1245   1977: contig of 733 bp in length
            *   1978   1987: gap of unknown length
            *   1988   2718: contig of 731 bp in length
            *   2719   2728: gap of unknown length
            *   2729   3253: contig of 525 bp in length
            *   3254   3263: gap of unknown length
            *   3264   3780: contig of 517 bp in length
            *   3781   3790: gap of unknown length
            *   3791   4504: contig of 714 bp in length
            *   4505   4514: gap of unknown length
            *   4515   5224: contig of 710 bp in length
            *   5225   5234: gap of unknown length
            *   5235   5734: contig of 500 bp in length
            *   5735   5744: gap of unknown length
            *   5745   6474: contig of 730 bp in length
            *   6475   6484: gap of unknown length
            *   6485   7191: contig of 707 bp in length
            *   7192   7201: gap of unknown length
            *   7202   7929: contig of 728 bp in length
            *   7930   7939: gap of unknown length
            *   7940   8667: contig of 728 bp in length
            *   8668   8677: gap of unknown length
            *   8678   8924: contig of 247 bp in length
            *   8925   8934: gap of unknown length
            *   8935   9657: contig of 723 bp in length
            *   9658   9667: gap of unknown length
            *   9668   10185: contig of 518 bp in length
            *   10186   10195: gap of unknown length
            *   10196   10711: contig of 516 bp in length
            *   10712   10721: gap of unknown length
            *   10722   11236: contig of 515 bp in length
            *   11237   11246: gap of unknown length
            *   11247   11759: contig of 513 bp in length
            *   11760   11769: gap of unknown length
            *   11770   12283: contig of 514 bp in length
            *   12284   12293: gap of unknown length
            *   12294   12616: contig of 323 bp in length
            *   12617   12626: gap of unknown length
            *   12627   13350: contig of 724 bp in length
            *   13351   13360: gap of unknown length
            *   13361   13877: contig of 517 bp in length
            *   13878   13887: gap of unknown length
            *   13888   14402: contig of 515 bp in length

            *   45715   gap of unknown length
            *   46634: contig of 920 bp in length
            *   46635   gap of unknown length
            *   47535: contig of 901 bp in length
            *   47536   gap of unknown length
            *   48461: contig of 926 bp in length
            *   48462   gap of unknown length
            *   49352: contig of 891 bp in length
            *   49353   gap of unknown length
            *   50255: contig of 903 bp in length
            *   50256   gap of unknown length
            *   51180: contig of 925 bp in length
            *   51181   gap of unknown length
            *   52074: contig of 894 bp in length
            *   52075   gap of unknown length
            *   52993: contig of 919 bp in length
            *   52994   gap of unknown length
            *   53936: contig of 943 bp in length
            *   53937   gap of unknown length
            *   54853: contig of 917 bp in length
            *   54854   gap of unknown length
            *   55782: contig of 929 bp in length
            *   55783   gap of unknown length
            *   56714: contig of 932 bp in length
            *   56715   gap of unknown length
            *   57619: contig of 905 bp in length
            *   57620   gap of unknown length
            *   58556: contig of 937 bp in length
            *   58557   gap of unknown length
            *   59455: contig of 899 bp in length
            *   59456   gap of unknown length
            *   60384: contig of 929 bp in length
            *   60385   gap of unknown length
            *   61280: contig of 896 bp in length
            *   61281   gap of unknown length
            *   62186: contig of 906 bp in length
            *   62187   gap of unknown length
            *   63081: contig of 895 bp in length
            *   63082   gap of unknown length
            *   63964: contig of 883 bp in length
            *   63965   gap of unknown length
            *   64876: contig of 912 bp in length
            *   64877   gap of unknown length
            *   65831: contig of 955 bp in length

Query Match   9.4%; Score 42.6; DB 54; Length 76968;
Best Local Similarity 51.3%;
Matches 122; Conservative 0; Mismatches 112; Indels 4; Gaps 1;

```
Qy    210 CCTCCAGGGGGGGACCTGGCTACTTGCCCCCGAGGCTT----CGCCGTCACCGGCTGCACT 265
          || || ||| ||  ||   |||||||| ||| |||||   || ||| |||| || ||
Db  24356 CCGCGCCGCCGGGCCCGCGGGCCGCCNCGCCCCCCGGGGCGCGNGGGCCCCCGGCGCCCG 24415

Qy    266 TGTGGCTCCGCCCTGTGGCTCGTGGGATGTGCGCGCCGAGACCACATGTCACTGCCAGTGC 325
          | | ||| | |||   ||| ||| | ||| |||  |||   ||| |||| | ||| |
Db  24416 GGGCCCCCGGCGGGGGCCCGCCCCCCGGGGGGGGCCCGGGGGGGCGGGGCGCG 24475

Qy    326 GCGGGCATGGACTGGACCGGAGCGGCGCTGCTGCTGTCGTGTCGTGTGCAGCCCTGAGGTCGCGCCCAG 385
          | | | ||| ||  |||   ||| | ||   |||| || |   |||| ||
Db  24476 CCGCGGCCCGGCGGGGGGCGCGGGGGGCCCGCGGGGGCCCCCGCCGGCCCGCCCC 24535

Qy    386 CGCCTGCACAGCCGGCGGGGCCGGAGCGGGCTCCAGGTCCGGAGGGGTTGCGGGGGAGCTGG 443
          |||| | | ||| ||  || ||||| || | |||| | |||| |||| ||
Db  24536 CGCGGCGGCCNCGGCGCGGGGCCCCCCGGCCGGGGGCGGGGCGGGGGCGGGGG 24593
```

```
*  14403  14412: gap of unknown length
*  14413  15117: contig of 705 bp in length
*  15118  15127: gap of unknown length
*  15128  16053: contig of 926 bp in length
*  16054  16063: gap of unknown length
*  16064  17174: contig of 1111 bp in length
*  17175  17184: gap of unknown length
*  17185  17917: contig of 733 bp in length
*  17918  17927: gap of unknown length
*  17928  18667: contig of 740 bp in length
*  18668  18677: gap of unknown length
*  18678  19426: contig of 749 bp in length
*  19427  19436: gap of unknown length
*  19437  20191: contig of 755 bp in length
*  20192  20201: gap of unknown length
*  20202  20926: contig of 725 bp in length
*  20927  20936: gap of unknown length
*  20937  21655: contig of 719 bp in length
*  21656  21665: gap of unknown length
*  21666  22400: contig of 735 bp in length
*  22401  22410: gap of unknown length
*  22411  22925: contig of 515 bp in length
*  22926  22935: gap of unknown length
*  22936  23450: contig of 515 bp in length
*  23451  23460: gap of unknown length
*  23461  23972: contig of 512 bp in length
*  23973  23982: gap of unknown length
*  23983  24722: contig of 740 bp in length
*  24723  24732: gap of unknown length
*  24733  25050: contig of 318 bp in length
*  25051  25060: gap of unknown length
*  25061  25765: contig of 705 bp in length
*  25766  25775: gap of unknown length
*  25776  26513: contig of 738 bp in length
*  26514  26523: gap of unknown length
*  26524  27260: contig of 737 bp in length
*  27261  27270: gap of unknown length
*  27271  27793: contig of 523 bp in length
*  27794  27803: gap of unknown length
*  27804  28538: contig of 735 bp in length
*  28539  28548: gap of unknown length
*  28549  29062: contig of 514 bp in length
*  29063  29072: gap of unknown length
*  29073  29585: contig of 513 bp in length
*  29586  29595: gap of unknown length
*  29596  30108: contig of 513 bp in length
*  30109  30118: gap of unknown length
*  30119  30634: contig of 516 bp in length
*  30635  30644: gap of unknown length
*  30645  31385: contig of 741 bp in length
*  31386  31395: gap of unknown length
*  31396  32384: contig of 989 bp in length
*  32385  32394: gap of unknown length
*  32395  33144: contig of 750 bp in length
*  33145  33154: gap of unknown length
*  33155  33891: contig of 735 bp in length
*  33892  33901: gap of unknown length
*  33902  34637: contig of 737 bp in length
*  34638  34647: gap of unknown length
*  34648  35160: contig of 736 bp in length
*  35161  35170: gap of unknown length
*  35171  35685: contig of 513 bp in length
*  35686  35695: gap of unknown length
*  35696  36127: contig of 515 bp in length
*  36128  36137: gap of unknown length
*  36138  36651: contig of 432 bp in length
*  36652  36661: gap of unknown length
*  36662  37395: contig of 514 bp in length
*  37396  37405: gap of unknown length
*  37406  38129: contig of 734 bp in length
*  38130  38139: gap of unknown length
*  38594  38603: contig of 724 bp in length

*  38604  39023: gap of unknown length
*  39024  39033: contig of 420 bp in length
*  39034  39549: gap of unknown length
*  39550  39559: contig of 516 bp in length
*  40309  40319: gap of unknown length
*  40320  41044: contig of 750 bp in length
*  41045  41054: gap of unknown length
*  41055  41572: contig of 725 bp in length
*  41573  41582: gap of unknown length
*  41583  42095: contig of 518 bp in length
*  42096  42105: gap of unknown length
*  42106  42619: contig of 513 bp in length
*  42620  42629: gap of unknown length
*  42630  43143: contig of 514 bp in length
*  43144  43153: gap of unknown length
*  43154  43651: contig of 514 bp in length
*  43652  43661: gap of unknown length
*  43662  44171: contig of 498 bp in length
*  44172  44181: gap of unknown length
*  44182  44918: contig of 510 bp in length
*  44919  44928: gap of unknown length
*  44929  45389: contig of 737 bp in length
*  45390  45399: gap of unknown length
*  45400  45917: contig of 461 bp in length
*  45918  45927: gap of unknown length
*  45928  46444: contig of 518 bp in length
*  46445  46454: gap of unknown length
*  46455  46875: contig of 517 bp in length
*  46876  46885: gap of unknown length
*  46886  47331: contig of 421 bp in length
*  47332  47341: gap of unknown length
*  47342  47735: contig of 446 bp in length
*  47736  47745: gap of unknown length
*  47746  48373: contig of 394 bp in length
*  48374  48383: gap of unknown length
*  48384  49073: contig of 628 bp in length
*  49074  49083: gap of unknown length
*  49084  49820: contig of 690 bp in length
*  49821  49830: gap of unknown length
*  49831  50566: contig of 737 bp in length
*  50567  50576: gap of unknown length
*  50577  51319: contig of 736 bp in length
*  51320  51329: gap of unknown length
*  51330  51845: contig of 743 bp in length
*  51846  51855: gap of unknown length
*  51856  52366: contig of 516 bp in length
*  52367  52376: gap of unknown length
*  52377  52890: contig of 511 bp in length
*                contig of 514 bp in length
```

```
Query Match            9.4%;   Score 42.6;   DB 44;   Length 164520;
Best Local Similarity 50.6%;   Pred. No. 3.4;
Matches 128;  Conservative 0;  Mismatches 124;  Indels 1;  Gaps 1;

Qy     188 GGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGGACCTGCTACTTGCCCCGAGGCTTC 247
           ||| | || |||| | ||| || |||| ||||||| |  || ||  || | ||||
Db 120880 GGCGGGCTCGCGGAGGGGCCCGGGGGGGGAGAGGGGCCGGGGTGGCCCCGCCAGGCAAA 120939

Qy     248 GCCGTCACCGGTTGCGCCCCGCCACTGTGGCTCCGCCTGTGCGCCGTCGGGATGTGGCC 307
            |||  || || |  |  |  ||| | || | | ||| | || |  ||| | |||
Db 120940 GCCGTTGCGGCCCCGCCGGTGCGCGTCGCAGTGGTGTGCGGGTGGCGGGTGGCGGCGGCC 120999

Qy     308 ACATGTCACTGCCAGTGCGGCGCATGGCGGCGAGCGCGTGCTGCTGTCGT-GTGCA 366
            | ||  |||  ||    | ||||  |  | | ||||  |  ||  ||||| | |||
Db 121000 CGCGGGGGTGAGTGCGGCCTCCCGCCGCCGCCTGAGGGGGGCGGCCGCGGGGGCGGGGGG 121059

Qy     367 GCCCTGAGGTCGCGCCAGCGCGTGCACAGACGCGGGCGAGGCGGCTCCAGGTCCGGAGG 426
            ||||  |  | ||  ||  | || || |  || |  |  | |  ||  |||| ||||
Db 121060 GCGGGGTGAGTGCGGCGGGGGCCTCGTGTTGGCGGCGGTGTCGGGGGGCGGGGGG 121119

Qy     427 GGTTGCGGGGGAG 439
           || | ||||||||
Db 121120 GGGCGGGGCGGGG 121132
```

```
RESULT 10
AC017082
LOCUS        AC017082    174707 bp    DNA        HTG        09-DEC-1999
DEFINITION   Homo sapiens clone RP11-472M4, LOW-PASS SEQUENCE SAMPLING.
ACCESSION    AC017082
VERSION      AC017082.1  GI:6554065
KEYWORDS     HTG; HTGS_PHASE0.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 174707)
AUTHORS      Waterston,R.H.
TITLE        The sequence of Homo sapiens clone
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 174707)
AUTHORS      Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (09-DEC-1999) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
COMMENT      Center project name: H_NH0567N19.
           * NOTE: This record contains 326 individual
           * sequencing reads that have not been assembled into
           * contigs. Runs of N are used to separate the reads
           * and the order in which they appear is completely
           * arbitrary. Low-pass sequence sampling is useful for
           * identifying clones that may be gene-rich and allows
           * overlap relationships among clones to be deduced.
           * However, it should not be assumed that this clone
           * will be sequenced to completion. In the event that
           * the record is updated, the accession number will
           * be preserved.
           *       1     386: contig of 386 bp in length
           *     387     396: gap of unknown length
           *     397     910: contig of 514 bp in length
           *     911     920: gap of unknown length
           *     921    1200: contig of 280 bp in length
           *    1201    1210: gap of unknown length
           *    1211    1722: contig of 512 bp in length
           *    1723    1732: gap of unknown length
           *    1733    2078: contig of 346 bp in length
           *    2079    2088: gap of unknown length
           *    2089    2600: contig of 512 bp in length
           *    2601    2610: gap of unknown length
           *    2611    3122: contig of 512 bp in length
           *    3123    3132: gap of unknown length
           *    3133    3643: contig of 511 bp in length
           *    3644    3653: gap of unknown length
           *    3654    4166: contig of 513 bp in length
           *    4167    4176: gap of unknown length
           *    4177    4689: contig of 513 bp in length
           *    4690    4699: gap of unknown length
           *    4700    5211: contig of 512 bp in length
           *    5212    5221: gap of unknown length
           *    5222    5666: contig of 445 bp in length
           *    5667    5676: gap of unknown length
           *    5677    6190: contig of 514 bp in length
           *    6191    6200: gap of unknown length
           *    6201    6714: contig of 514 bp in length
           *    6715    6724: gap of unknown length
           *    6725    7236: contig of 512 bp in length
           *    7237    7246: gap of unknown length
           *    7247    7729: contig of 483 bp in length
           *    7730    7739: gap of unknown length
           *    7740    8591: contig of 852 bp in length
           *    8592    8601: gap of unknown length
           *    8602    9113: contig of 512 bp in length
           *    9114    9123: gap of unknown length
           *    9124    9634: contig of 511 bp in length
           *    9635    9644: gap of unknown length
           *    9645   10156: contig of 512 bp in length
           *   10157   10166: gap of unknown length
           *   10167   10670: contig of 504 bp in length
           *   10671   10680: gap of unknown length
           *   10681   10917: contig of 237 bp in length
           *   10918   10927: gap of unknown length
           *   10928   11132: contig of 205 bp in length
           *   11133   11142: gap of unknown length
           *   11143   11648: contig of 506 bp in length
           *   11649   11658: gap of unknown length
           *   11659   12170: contig of 512 bp in length
           *   12171   12180: gap of unknown length
           *   12181   13038: contig of 858 bp in length
           *   13039   13048: gap of unknown length
           *   13049   13472: contig of 424 bp in length
           *   13473   13482: gap of unknown length
           *   13483   13994: contig of 512 bp in length
           *   13995   14004: gap of unknown length
           *   14005   14436: contig of 432 bp in length
           *   14437   14446: gap of unknown length
           *   14447   14955: contig of 509 bp in length
           *   14956   14965: gap of unknown length
           *   14966   15478: contig of 512 bp in length
           *   15478   15487: gap of unknown length
           *   15488   15999: contig of 512 bp in length
           *   16000   16009: gap of unknown length
           *   16010   16397: contig of 388 bp in length
           *   16398   16407: gap of unknown length
           *   16408   16863: contig of 456 bp in length
           *   16864   16873: gap of unknown length
           *   16874   17385: contig of 512 bp in length
           *   17386   17395: gap of unknown length
           *   17396   18220: contig of 825 bp in length
           *   18221   18230: gap of unknown length
           *   18231   18742: contig of 512 bp in length
           *   18743   18752: gap of unknown length
           *   18753   18876: contig of 124 bp in length
           *   18877   18886: gap of unknown length
           *   18887   19400: contig of 514 bp in length
           *   19401   19410: gap of unknown length
           *   19411   19573: contig of 163 bp in length
           *   19574   19583: gap of unknown length
           *   19584   20429: contig of 846 bp in length
           *   20430   20439: gap of unknown length
           *   20440   21291: contig of 852 bp in length
           *   21292   21301: gap of unknown length
           *   21302   21775: contig of 474 bp in length
           *   21776   21785: gap of unknown length
           *   21786   22296: contig of 511 bp in length
           *   22297   22306: gap of unknown length
           *   22307   22810: contig of 504 bp in length
           *   22811   22820: gap of unknown length
           *   22821   23332: contig of 512 bp in length
           *   23333   23342: gap of unknown length
           *   23343   23854: contig of 512 bp in length
           *   23855   23864: gap of unknown length
           *   23865   24376: contig of 512 bp in length
           *   24377   24386: gap of unknown length
           *   24387   24900: contig of 514 bp in length
           *   24901   24910: gap of unknown length
           *   24911   25424: contig of 514 bp in length
           *   25425   25434: gap of unknown length
           *   25435   25828: contig of 394 bp in length
           *   25829   25838: gap of unknown length
           *   25839   26350: contig of 512 bp in length
           *   26351   26360: gap of unknown length
           *   26361   26871: contig of 511 bp in length
           *   26872   26881: gap of unknown length
           *   26882   28779: contig of 1898 bp in length
           *   28780   28789: gap of unknown length
           *   28790   29300: contig of 511 bp in length
           *   29301   29310: gap of unknown length
           *   29311   29774: contig of 464 bp in length
```

```
Db  99126  GGGCGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 99185

Qy    394  CAGGCGGCGGCGGAGCGGCGGCTCCAGTCCGGAGGGTTGCGGGGGAGCTGAAATAA 449
           ||||| |||    ||  ||   |||| ||||||| ||||||||||||   ||||| ||
Db  99186  GGGCGGCCCGGACGGCCCGGAGGGGAGGGGGGCCGAGAGGAGAGAGAGAGAGAGAAA 99241
```

```
RESULT 11
AC023395/c
LOCUS       AC023395       76199 bp    DNA                    HTG       14-FEB-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-484N16 map 18, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC023395
VERSION     AC023395.1  GI:6970549
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 76199)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-484N16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 76199)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigillo,J., Vassillev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----------------------  Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ----------------------  Project Information
                Center project name: L5583
                Center clone name: 484_N_16
            ----------------------
            * NOTE: This record contains 89 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *
            *          1  848: contig of 848 bp in length
            *             gap of unknown length
```

```
*  29775  29784: gap of unknown length
*  29785  30727: contig of 943 bp in length
*  30728  30737: gap of unknown length
*  31238  31251: contig of 514 bp in length
*  31252  31261: gap of unknown length
*  31262  31773: contig of 512 bp in length
*  31774  31783: gap of unknown length
*  31784  32199: contig of 416 bp in length
*  32200  32209: gap of unknown length
*  32210  32721: contig of 512 bp in length
*  32722  32731: gap of unknown length
*  32732  32977: contig of 246 bp in length
*  32978  32987: gap of unknown length
*  32988  33467: contig of 480 bp in length
*  33468  33477: gap of unknown length
*  33478  33991: contig of 514 bp in length
*  33992  34001: gap of unknown length
*  34002  34514: contig of 513 bp in length
*  34515  34524: gap of unknown length
*  34525  35373: contig of 849 bp in length
*  35374  35383: gap of unknown length
*  35384  35895: contig of 512 bp in length
*  35896  35905: gap of unknown length
*  35906  36339: contig of 434 bp in length
*  36340  36349: gap of unknown length
*  36350  37203: contig of 854 bp in length
*  37204  37213: gap of unknown length
*  37214  37427: contig of 214 bp in length
*  37428  37437: gap of unknown length
*  37438  37949: contig of 512 bp in length
*  37950  37959: gap of unknown length
*  37960  38829: contig of 870 bp in length
*  38830  38839: gap of unknown length
*  38840  39343: contig of 504 bp in length
*  39344  39353: gap of unknown length
*  39354  39670: contig of 317 bp in length
*  39671  39680: gap of unknown length
*  39681  40193: contig of 513 bp in length
*  40194  40203: gap of unknown length
*  40204  40524: contig of 321 bp in length
*  40525  40534: gap of unknown length
*  40535  41048: contig of 514 bp in length
*  41049  41058: gap of unknown length
*  41059  41141: contig of 83 bp in length
*  41142  41151: gap of unknown length
*  41152  41664: contig of 513 bp in length
*  41665  41674: gap of unknown length
*  41675  42171: contig of 497 bp in length
*  42172  42181: gap of unknown length
*  42182  42695: contig of 514 bp in length
*  42696  42705: gap of unknown length
*  42706  43219: contig of 514 bp in length
*  43220  43229: gap of unknown length
*  43230  43742: contig of 513 bp in length
*  43743  43752: gap of unknown length
*  43753  44260: contig of 508 bp in length
*  44261  44270: gap of unknown length
*  44271  44784: contig of 514 bp in length
```

```
Query Match             9.4%;  Score 42.4;  DB 43;  Length 174707;
Best Local Similarity  48.7%;  Pred. No. 3.7;
Matches  115;  Conservative   0;  Mismatches  121;  Indels   0;  Gaps   0;

Qy    214  CAGGGGGACCTGCTACTTGCCCCCGAGGCTTCGCCCTCACCGGCTGCACTTGTGGCTC 273
           ||||||||  |                     ||||||||||||||||||||||||||
Db  99006  CGGGGGGAGCTGCGGGGGGGCCGGGCCGGGGCCGGGGGGCCGGGGGAGGGGGGGGGGG 99065

Qy    274  CGCCTGCTCGTCGGGATGTGCCGGCCGAGACCACATGTCACTGCGGCGGCCGGGCAT 333
           |||||||| |  |           | | |      || |||||||||| |||||||||
Db  99066  CGGCGGGGCGGGGGGGGCCGGGCGAGGGGGCGGCGGAGTGGCCGGGGGCCGGGGGGGC 99125

Qy    334  GGACTGGACCCGAGCCGGCGCTGCTCGTCGTGCAGCCCTGAGGTCGCGGCGCGCGTGCA 393
           ||  ||| |||  ||  ||   |||| ||  ||| ||||||||| || ||||||||||||
```

```
*    849 contig of 875 bp in length
*   1723 gap of unknown length
*   1724 contig of 844 bp in length
*   2567 gap of unknown length
*   2568 contig of 855 bp in length
*   3422 gap of unknown length
*   3423 contig of 864 bp in length
*   4286 gap of unknown length
*   4287 contig of 903 bp in length
*   5189 gap of unknown length
*   5190 contig of 881 bp in length
*   6070 gap of unknown length
*   6071 contig of 877 bp in length
*   6947 gap of unknown length
*   6948 contig of 856 bp in length
*   7803 gap of unknown length
*   7804 contig of 859 bp in length
*   8662 gap of unknown length
*   8663 contig of 860 bp in length
*   9522 gap of unknown length
*   9523 contig of 896 bp in length
*  10418 gap of unknown length
*  10419 contig of 860 bp in length
*  11278 gap of unknown length
*  11279 contig of 855 bp in length
*  12133 gap of unknown length
*  12134 contig of 836 bp in length
*  12969 gap of unknown length
*  12970 contig of 840 bp in length
*  13809 gap of unknown length
*  13810 contig of 872 bp in length
*  14681 gap of unknown length
*  14682 contig of 856 bp in length
*  15537 gap of unknown length
*  15538 contig of 852 bp in length
*  16389 gap of unknown length
*  16390 contig of 842 bp in length
*  17231 gap of unknown length
*  17232 contig of 838 bp in length
*  18069 gap of unknown length
*  18070 contig of 857 bp in length
*  18926 gap of unknown length
*  18927 contig of 854 bp in length
*  19780 gap of unknown length
*  19781 contig of 834 bp in length
*  20614 gap of unknown length
*  20615 contig of 824 bp in length
*  21438 gap of unknown length
*  21439 contig of 861 bp in length
*  22299 gap of unknown length
*  22300 contig of 874 bp in length
*  23173 gap of unknown length
*  23174 contig of 836 bp in length
*  24009 gap of unknown length
*  24010 contig of 870 bp in length
*  24879 gap of unknown length
*  24880 contig of 891 bp in length
*  25770 gap of unknown length
*  25771 contig of 884 bp in length
*  26654 gap of unknown length
*  26655 contig of 854 bp in length
*  27508 gap of unknown length
*  27509 contig of 867 bp in length
*  28375 gap of unknown length
*  28376 contig of 865 bp in length
*  29240 gap of unknown length
*  29241 contig of 864 bp in length
*  30104 gap of unknown length
*  30105 contig of 839 bp in length
*  30943 gap of unknown length
*  30944 contig of 839 bp in length
*  31782 gap of unknown length
*  31783 contig of 845 bp in length
*  32627 gap of unknown length

*  32628 gap of unknown length
*  33465 contig of 838 bp in length
*  33466 gap of unknown length
*  34329 contig of 864 bp in length
*  34330 gap of unknown length
*  35214 contig of 885 bp in length
*  35215 gap of unknown length
*  36040 contig of 826 bp in length
*  36041 gap of unknown length
*  36890 contig of 850 bp in length
*  36891 gap of unknown length
*  37721 contig of 831 bp in length
*  37722 gap of unknown length
*  38573 contig of 852 bp in length
*  38574 gap of unknown length
*  39417 contig of 844 bp in length
*  39418 gap of unknown length
*  40256 contig of 839 bp in length
*  40257 gap of unknown length
*  41130 contig of 874 bp in length
*  41131 gap of unknown length
*  41993 contig of 863 bp in length
*  41994 gap of unknown length
*  42840 contig of 847 bp in length
*  42841 gap of unknown length
*  43715 contig of 875 bp in length
*  43716 gap of unknown length
*  44582 contig of 867 bp in length
*  44583 gap of unknown length
*  45460 contig of 878 bp in length
*  45461 gap of unknown length
*  46282 contig of 822 bp in length
*  46283 gap of unknown length
*  47130 contig of 848 bp in length
*  47131 gap of unknown length
*  47970 contig of 840 bp in length
*  47971 gap of unknown length
*  48796 contig of 826 bp in length
*  48797 gap of unknown length
*  49613 contig of 817 bp in length
*  49614 gap of unknown length
*  50472 contig of 859 bp in length
*  50473 gap of unknown length
*  51325 contig of 853 bp in length
*  51326 gap of unknown length
*  52222 contig of 897 bp in length
*  52223 gap of unknown length
*  53066 contig of 844 bp in length
*  53067 gap of unknown length
*  53932 contig of 866 bp in length
*  53933 gap of unknown length
*  54805 contig of 873 bp in length
*  54806 gap of unknown length
*  55651 contig of 846 bp in length
*  55652 gap of unknown length
*  56494 contig of 843 bp in length
*  56495 gap of unknown length
*  57356 contig of 862 bp in length
*  57357 gap of unknown length
*  58238 contig of 882 bp in length
*  58239 gap of unknown length
*  59091 contig of 853 bp in length
```

```
Query Match          9.2%;   Score 41.8;  DB 55;  Length 76199;
Best Local Similarity 48.58%;  Pred. No. 5.5;
Matches 115; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

Qy   216 GGGGGGACTGGCTACTTGCCCCGAGGCTTCGCCGTCACCGGCTGCACTTGTGGCTCCG 275
         |-||||-|||-||--||-|-|||-|| - |||-| |-|||- | -|||||-
Db 69824 GCGGGGCTGCGGCCTCGCGGGCTCTGCGGCCCGGTCCCTCGGCACCTCCGGCGGCTCGG 69765
Qy   276 CCTGTGGCTCGTGGGATGTGCGCGCCGAGACACCATGTCACTGCCAGTGCGGCGGCATGG 335
```

```
Db  69764 GCTCGGCTCGGGGTCGGGCGGCGGCGGACCCGCCGCAGGCGGCGGCCCTGGGGGCCG 69705
                ||  |||||| ||   ||||| |||| ||| ||  ||  ||| |||
Qy    336 ACTGGACCGGAGCGGCCTGCTGTCGTGTGCAGCCCTGAGGTCGCGCAGCGGCTGCACA 395
                |  ||  | |||  |  || | |||||  |  || ||||||||
Db  69704 GCGCTGGGACTCGGGGACGCCCCCTCGCCCACCCGGAGGTCGCGGCCTGGTTGGAG 69645
           | ||| |||  ||| || |||   ||| ||| ||   ||||    |||
Qy    396 GCGCGGCGGAGGCGGCGGCGCTCCAGGTCCGAGGGGTTCGGGGGAGCTGGAAATAAACC 452
           | ||| |||||  || ||| |  |  ||| || || |||     ||||
Db  69644 GTGCAGGCCGGGGCGGCGTGCTCTCGGCGGCGTGAAGGTGCCGGGCGGCCCGGCTGCGCCACTCACC 69588
```

RESULT 12
AC023660
LOCUS   AC023660   81213 bp   DNA   HTG   16-FEB-2000
DEFINITION Homo sapiens chromosome 12 clone RP11-400B24 map 12, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC023660
VERSION   AC023660.1 GI:6980354
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 81213)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE   Homo sapiens chromosome 12, clone RP11-400B24
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 81213)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkly,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKeown,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pollera,V., Pisani,C., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (16-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: L6752
            Center clone name: 400_B_24
            ----
            * NOTE: This record contains 87 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *   1   927: contig of 927 bp in length
            *   928   1864: gap of unknown length
            *   1865   2823: contig of 937 bp in length
            *   2824   3737: gap of unknown length
            *   3738   4666: contig of 959 bp in length
            *   4667   5618: gap of unknown length
            *   5619   6565: contig of 947 bp in length
            *   6566   7485: gap of unknown length
            *   7486   8411: contig of 926 bp in length
            *   8412   9361: gap of unknown length
            *   9362 10300: contig of 950 bp in length
            *   10301 11237: gap of unknown length
            *   11238 12202: contig of 965 bp in length
            *   12203 13143: gap of unknown length
            *   13144 14043: contig of 941 bp in length
            *   14044 14956: gap of unknown length
            *   14957 15891: contig of 935 bp in length
            *   15892 16845: gap of unknown length
            *   16846 17797: contig of 952 bp in length
            *   17798 18735: gap of unknown length
            *   18736 19632: contig of 938 bp in length
            *   19633 20306: gap of unknown length
            *   20307 21221: contig of 915 bp in length
            *   21222 22167: gap of unknown length
            *   22168 23100: contig of 933 bp in length
            *   23101 24065: gap of unknown length
            *   24066 25005: contig of 940 bp in length
            *   25006 25897: gap of unknown length
            *   25898 26810: contig of 913 bp in length
            *   26811 27775: gap of unknown length
            *   27776 28715: contig of 940 bp in length
            *   28716 29643: gap of unknown length
            *   29644 30592: contig of 949 bp in length
            *   30593 31515: gap of unknown length
            *   31516 32443: contig of 928 bp in length
            *   gap of unknown length
```

us-09-099-898-1.rge

Fri May 26 10:31:06 2000

```
Qy    279 GTGGCTCGTGGGATGTGCGCCCGAGACCACATGTCACTGCCAGTGCGGCGGCATGGACT 338
          |||  ||||| |||   |||       |||-|-|-||-| ||| |||| |||| |||
Db  25644 GCGGCGCGGGGCGCGGGGGCGGCCGGTTGGGGGGGTCGCGGGGGGGGGGGGCTGGGCG 25703

Qy    339 GGACCCGAGACGCGCGCTCTGTCTGTGCAGCCCTGAGTGTCGGGAGTCGCGGTCCACAGCG 398
           |||         ||||  | |-||||||| |||   ||||  ||||
Db  25704 GCGGGCGCGTGGGCGCGGGGCGTCGGGGGGGTCGGGGGGGGGGGGGGGGCGGGCGCGCCG 25763

Qy    399 CGGGCGGAGGCGGCTCCAGTTCCGGAGGGGTTGCGGGGGACTGG 443
           ||||||||||  | |||||-||| ||||||||| ||| ||-||
Db  25764 TGGTGTCGCGGGGCGGCGGCGGGGCGTCGGGGGGGTGCGGGGGGGG 25808
```

RESULT 13
AC005776/c
LOCUS   AC005776   40883 bp   DNA   PRI   06-OCT-1998
DEFINITION   Homo sapiens chromosome 19, cosmid R28550, complete sequence.
ACCESSION   AC005776
VERSION   AC005776.1 GI:3702305
KEYWORDS   HTG.
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 40883)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Skowronski,E., Viswanathan,V.,
            Burkhart-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Stilwagen,S.,
            Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
            Danganan,L., Erler,A., Christensen,M., Georgescu,A., Avila,J.,
            Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
            Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
            Krommiller,B., Arellano,A., Saunders,C., Ow,D., Nolan,M., Trong,S.,
            Kobayashi,A., Olsen,A.S. and Carrano,A.V.
  TITLE   Sequence analysis of a 1.5 Mb OLFR-rich region in 19p13.1
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 40883)
  AUTHORS   Lamerdin,J.E.
  TITLE   Direct Submission
  JOURNAL   Submitted (05-OCT-1998) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT   Map and sequence oriented from p telomere to centromere. Cosmid
            R28550 overlaps cosmid R28194 (AC003111) to the left from bases 1
            to 504 of this accession, and is separated from cosmid R33907 to
            the right by a sequence gap of <2 kb. Additional chromosome 19 map
            and sequence information are available at:
            http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES   Location/Qualifiers
     source   1..40883
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="R28550"
                     /chromosome="19"
                     /map="19p13.1 between OLFR and D19S885"
                     /cell_line="5HL2-B"
                     /clone_lib="LL19NC02 F chromosome 19 specific cosmid
                     library"
                     /note="Cosmid library constructed at LLNL from flow-sorted
                     chromosomes from human-hamster hybrid 5HL2-B, which
                     carries chromosome 19 as its only human chromosome."
     repeat_region   1..245
                     /rpt_family="AluSq"
     repeat_region   246..442
                     /rpt_family="L1MB8"
     repeat_region   443..754
                     /rpt_family="AluSq"
     repeat_region   770..963
                     /rpt_family="L1MB1"
     repeat_region   1161..1282
                     /rpt_family="AluSq/x"
     repeat_region   1286..1585
                     /rpt_family="AluSq"
     repeat_region   1588..1880
                     /rpt_family="AluSx"
```

```
*  32444    33356: contig of 913 bp in length
*            gap of unknown length
*  33357    34245: contig of 889 bp in length
*            gap of unknown length
*  34246    35131: contig of 886 bp in length
*            gap of unknown length
*  35132    36058: contig of 927 bp in length
*            gap of unknown length
*  36059    37021: contig of 963 bp in length
*            gap of unknown length
*  37022    37973: contig of 952 bp in length
*            gap of unknown length
*  37974    38897: contig of 924 bp in length
*            gap of unknown length
*  38898    39859: contig of 962 bp in length
*            gap of unknown length
*  39860    40805: contig of 946 bp in length
*            gap of unknown length
*  40806    41761: contig of 956 bp in length
*            gap of unknown length
*  41762    42694: contig of 933 bp in length
*            gap of unknown length
*  42695    43653: contig of 959 bp in length
*            gap of unknown length
*  43654    44591: contig of 938 bp in length
*            gap of unknown length
*  44592    45529: contig of 938 bp in length
*            gap of unknown length
*  45530    46488: contig of 959 bp in length
*            gap of unknown length
*  46489    47418: contig of 930 bp in length
*            gap of unknown length
*  47419    48356: contig of 938 bp in length
*            gap of unknown length
*  48357    49317: contig of 961 bp in length
*            gap of unknown length
*  49318    50264: contig of 947 bp in length
*            gap of unknown length
*  50265    51229: contig of 965 bp in length
*            gap of unknown length
*  51230    52171: contig of 942 bp in length
*            gap of unknown length
*  52172    53094: contig of 923 bp in length
*            gap of unknown length
*  53095    54029: contig of 935 bp in length
*            gap of unknown length
*  54030    54949: contig of 920 bp in length
*            gap of unknown length
*  54950    55873: contig of 924 bp in length
*            gap of unknown length
*  55874    56819: contig of 946 bp in length
*            gap of unknown length
*  56820    57771: contig of 952 bp in length
*            gap of unknown length
*  57772    58738: contig of 967 bp in length
*            gap of unknown length
*  58739    59683: contig of 945 bp in length
*            gap of unknown length
*  59684    60608: contig of 925 bp in length
*            gap of unknown length
*  60609    61575: contig of 967 bp in length
*            gap of unknown length
*  61576    62526: contig of 951 bp in length
*            gap of unknown length
*  62527    63481: contig of 955 bp in length
*            gap of unknown length
*  63482    64431: contig of 950 bp in length
*            gap of unknown length
```

```
Query Match        9.2%;  Score 41.8;  DB 55;  Length 81213;
Best Local Similarity  53.3%;  Pred. No. 5.5;
Matches 88;  Conservative  0;  Mismatches 77;  Indels  0;  Gaps  0;
```

```
repeat_region    1891..2012
                 /rpt_family="AluJb"
repeat_region    2024..2326
                 /rpt_family="AluY"
repeat_region    2327..2616
                 /rpt_family="AluSq"
repeat_region    2776..2841
                 /rpt_family="LINE2"
repeat_region    3495..3622
                 /rpt_family="LINE2"
repeat_region    complement(3667..3766)
                 /rpt_family="MER3"
misc_feature     complement(4287..4396)
                 /note="predicted exon, program: grail2exons_human_1.3,
                 frame: 2, quality: good, score: 52.000"
repeat_region    4507..4787
                 /rpt_family="AluSq"
repeat_region    4808..5109
                 /rpt_family="AluY"
repeat_region    5283..5545
                 /rpt_family="AluSq"
repeat_region    6064..6207
                 /rpt_family="MIR"
repeat_region    6537..6577
                 /rpt_family="(TGG)n"
repeat_region    8122..8801
                 /rpt_family="AluJb"
repeat_region    8402..8565
                 /rpt_family="AluJo/FRAM"
repeat_region    8585..8889
                 /rpt_family="AluJb"
repeat_region    9886..10177
                 /rpt_family="AluSx"
repeat_region    10179..10251
                 /rpt_family="(GAAA)n"
repeat_region    complement(11205..11323)
                 /rpt_family="AluY"
repeat_region    11363..11494
                 /rpt_family="AluY"
repeat_region    11506..11820
                 /rpt_family="AluY"
repeat_region    11821..12003
                 /rpt_family="AluYa5"
repeat_region    12022..12128
                 /rpt_family="FLAM_A"
repeat_region    complement(12200..12259)
                 /rpt_family="MIR"
repeat_region    complement(12910..13002)
                 /rpt_family="AT_rich"
repeat_region    13371..14036
                 /rpt_family="MIR"
misc_feature     complement(14095..14209)
                 /note="predicted exon, program: grail2exons_human_1.3,
                 frame: 1, quality: excellent, score: 83.000"
repeat_region    14662..14951
                 /rpt_family="MIR"
repeat_region    complement(15085..15108)
                 /rpt_family="AluSp"
repeat_region    15109..15431
                 /rpt_family="AluSx"
repeat_region    15562..15770
                 /rpt_family="AluSx"
misc_feature     16107..16190
                 /note="predicted exon, program: grail2exons_human_1.3,
                 frame: 1, quality: excellent, score: 80.000"
repeat_region    16191..16497
                 /rpt_family="L1MB8"
repeat_region    16498..16937
                 /rpt_family="L1MB8"
repeat_region    complement(17602..17708)
                 /rpt_family="LINE2"
repeat_region    complement(17709..18005)
                 /rpt_family="AluSp"

repeat_region    18393..18690
                 /rpt_family="AluSc"
misc_feature     18706..18968
                 /note="BLASTN similarity to T41380 (105..368); match:
                 0.99; score: 3.8e100; database searched: est;
                 phlb12_19/ITV Homo sapiens cDNA clone phlb12_19/ITV."
repeat_region    19418..19703
                 /rpt_family="AluSx"
repeat_region    complement(20667..20801)
                 /rpt_family="GC_rich"
repeat_region    complement(20802..20829)
                 /rpt_family="(GA)n"
repeat_region    complement(20871..20955)
                 /rpt_family="(CGG)n"
repeat_region    complement(20982..21134)
                 /rpt_family="GC_rich"
repeat_region    complement(21237..21352)
                 /rpt_family="(CGG)n"
repeat_region    complement(21349..21470)
                 /rpt_family="(CGG)n"
repeat_region    21482..21559
                 /rpt_family="(CG)n"
repeat_region    complement(22010..22285)
                 /rpt_family="AluJo"
repeat_region    22319..22464
                 /rpt_family="AluSx"
repeat_region    22465..22757
                 /rpt_family="AluY"
repeat_region    22758..22935
                 /rpt_family="AluSx"
repeat_region    complement(23192..23484)
                 /rpt_family="AluJo"
repeat_region    23530..23829
                 /rpt_family="AluSx"
repeat_region    23866..24168
                 /rpt_family="AluSx"
repeat_region    complement(24456..24584)
                 /rpt_family="FRAM"
repeat_region    24597..24734
                 /rpt_family="AluSx"
repeat_region    24737..25071
                 /rpt_family="AluSq"
repeat_region    25072..25238
                 /rpt_family="AluSx"
repeat_region    complement(25239..25317)
                 /rpt_family="AluJb"
repeat_region    25677..25969
                 /rpt_family="AluY"
repeat_region    complement(26347..26643)
                 /rpt_family="AluSx"
repeat_region    complement(26647..26831)
                 /rpt_family="AluJo"
repeat_region    complement(26839..27000)
                 /rpt_family="FAM"
repeat_region    complement(27094..27394)
                 /rpt_family="AluSx"
repeat_region    complement(27415..27507)
                 /rpt_family="7SLRNA"
repeat_region    27599..27892
                 /rpt_family="AluSx"
misc_feature     27924..28121
                 /note="BLASTX similarity to 619788 (14..79); match: 0.5;
                 score: 6.7e17; database searched: nr; (L38826) L21
                 ribosomal protein [Homo sapiens]"
misc_feature     28037..28147
                 /note="BLASTX similarity to 619788 (52..88); match: 0.81;
                 score: 7.0e-15; database searched: nr; (L38826) L21
                 ribosomal protein [Homo sapiens]"
repeat_region    28159..28393

Query Match
Best Local Similarity    9.2%;        Score 41.6;   DB 11;   Length 40883;
                         45.2%;       Pred. No. 6.6;
Matches 152;  Conservative 0;  Mismatches 184;  Indels 0;  Gaps 0;
```

```
Qy      108  TGTGCTCCATGGAAGAAGCCATCAATGAGAGAGATCCAGGAGGTCGCCGGCTCCCTAATAT 167
             || || || |  |||| ||| |||||| ||| |||||  |||| |||| |  |
Db    21212  TGCCCTTTGGGGCGGCGGGGACGACACTGGGCGGCTGGCGGAGGCTCGCCGGCTCCGCG 21153

Qy      168  TAAGGCAATAAGCAGCATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGGACCTGG 227
             |||| ||  | ||||  ||||| |  ||||||||  | |||| |  | ||
Db    21152  TAACCTGCATTGCGTAATGGGCGGCGGGCGGGCCGCCACCTACGCCCGCGGCGGCGGGG 21093

Qy      228  CTACTTGCCCCCGAGGCTTCGCCGCTCACCCGGCTGCACTTGTGGCTCCGCTGTTGGCTCGT 287
             | |||| |||| |  ||| ||  || |||| | |||||  || | || ||| |||| |
Db    21092  GGCGGGGAGGCCGGGCTACGCGGGCTCGGGCCTCGGGCCTGCGGGGCCTCCCGGCCGCGTTCGTTCCG 21033

Qy      288  GGGATGTCGCGCGCGAGACCACATGTCACTGCCCAGTGCCGGGGCATGGACCGGAG 347
             || | ||| ||||  ||| |   | |  | |  || |   || |  | |||
Db    21032  GGCGGGGGCGGGGCCTGCGCCGTTGGTGTCGGGGCCGGCGGGGTCGGTTCGCGGCCGTGTCCTGCC 20973

Qy      348  CCGGCTGCTGTGTGTGCCCTGAGGTCGCCGGCGCCAGCCGGTGCACAGCGCGGGCGGAG 407
             ||| | ||  |  | |  |  ||  || || |  |||  ||| || ||  | ||| |
Db    20972  TTTGGTGCCGTCGGACCCCGGCGCTGGGCGTGGCCCCGCCGCGGGGGGTGGGGGGCCGGGC 20913

Qy      408  GCGGCTCCAGGTCCGGAGGGGTTGCGGGGGAGCTGG 443
             ||  || | |||||| |   | ||  |||  ||| ||
Db    20912  GCCGGCGGGGGGGGCCGGCGGGTCGGGGGGGCCGGGGG 20877


RESULT 14
AC023214/c
LOCUS       AC023214        68523 bp        DNA            HTG        09-FEB-2000
DEFINITION  Homo sapiens clone RP11-313B12, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC023214
VERSION     AC023214.1    GI:6957756
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 68523)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-313B12
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 68523)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,R., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            ----------- Project Information
              Center project name: L6599
              Center clone name: 313_B_12
            ----------- 
* NOTE: This record contains 88 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*          1       786:  contig of 786 bp in length
*        787      1549:  gap of unknown length
*       1550      2334:  contig of 785 bp in length
*       2335      3104:  gap of unknown length
*       3105      3880:  contig of 770 bp in length
*       3881      4652:  gap of unknown length
*       4653      5422:  contig of 776 bp in length
*       5423      6188:  gap of unknown length
*       6189      6969:  contig of 766 bp in length
*       6970      7738:  gap of unknown length
*       7739      8506:  contig of 769 bp in length
*       8507      9260:  gap of unknown length
*       9261     10044:  contig of 768 bp in length
*      10045     10826:  gap of unknown length
*      10827     11607:  contig of 784 bp in length
*      11608     12400:  gap of unknown length
*      12401     13166:  contig of 782 bp in length
*      13167     13934:  gap of unknown length
*      13935     14707:  contig of 781 bp in length
*      14708     15470:  gap of unknown length
*      15471     16330:  contig of 793 bp in length
*      16331     17102:  gap of unknown length
*      17103     17865:  contig of 766 bp in length
*      17866     18645:  gap of unknown length
*      18646     19405:  contig of 768 bp in length
*      19406     20202:  gap of unknown length
*      20203     20999:  contig of 773 bp in length
*      21000     21778:  gap of unknown length
*      21779     22546:  contig of 763 bp in length
*      22547     23318:  gap of unknown length
*      23319     24106:  contig of 860 bp in length
```

```
*  24107   24896: gap of unknown length
*                 contig of 790 bp in length
*  24897   25670: gap of unknown length
*                 contig of 774 bp in length
*  25671   26434: contig of 764 bp in length
*                 gap of unknown length
*  26435   27198: contig of 764 bp in length
*                 gap of unknown length
*  27199   27981: contig of 783 bp in length
*                 gap of unknown length
*  27982   28766: contig of 785 bp in length
*                 gap of unknown length
*  28767   29562: contig of 796 bp in length
*                 gap of unknown length
*  29563   30377: contig of 815 bp in length
*                 gap of unknown length
*  30378   31157: contig of 780 bp in length
*                 gap of unknown length
*  31158   31937: contig of 780 bp in length
*                 gap of unknown length
*  31938   32717: contig of 780 bp in length
*                 gap of unknown length
*  32718   33476: contig of 759 bp in length
*                 gap of unknown length
*  33477   34237: contig of 761 bp in length
*                 gap of unknown length
*  34238   34999: contig of 762 bp in length
*                 gap of unknown length
*  35000   35755: contig of 756 bp in length
*                 gap of unknown length
*  35756   36513: contig of 758 bp in length
*                 gap of unknown length
*  36514   37262: contig of 749 bp in length
*                 gap of unknown length
*  37263   38033: contig of 771 bp in length
*                 gap of unknown length
*  38034   38803: contig of 770 bp in length
*                 gap of unknown length
*  38804   39592: contig of 789 bp in length
*                 gap of unknown length
*  39593   40368: contig of 776 bp in length
*                 gap of unknown length
*  40369   41140: contig of 772 bp in length
*                 gap of unknown length
*  41141   41908: contig of 768 bp in length
*                 gap of unknown length
*  41909   42655: contig of 747 bp in length
*                 gap of unknown length
*  42656   43411: contig of 756 bp in length
*                 gap of unknown length
*  43412   44174: contig of 763 bp in length
*                 gap of unknown length
*  44175   44955: contig of 781 bp in length
*                 gap of unknown length
*  44956   45744: contig of 789 bp in length
*                 gap of unknown length
*  45745   46503: contig of 759 bp in length
*                 gap of unknown length
*  46504   47289: contig of 786 bp in length
*                 gap of unknown length
*  47290   48044: contig of 755 bp in length
*                 gap of unknown length
*  48045   48810: contig of 766 bp in length
*                 gap of unknown length
*  48811   49576: contig of 766 bp in length
*                 gap of unknown length
*  49577   50329: contig of 753 bp in length
*                 gap of unknown length
*  50330   51113: contig of 784 bp in length
*                 gap of unknown length
*  51114   51909: contig of 796 bp in length
*                 gap of unknown length
```

```
*  51910   52750: contig of 841 bp in length
*                 gap of unknown length
*  52751   53657: contig of 907 bp in length
*                 gap of unknown length
*  53658   54447: contig of 790 bp in length
*                 gap of unknown length
*                 gap of unknown length
*  54448   55213: contig of 766 bp in length
*                 gap of unknown length
*  55214   55990: contig of 777 bp in length
```

```
Query Match           9.1%;   Score 41.2;  DB 55;  Length 68523;
Best Local Similarity 49.1%;  Pred. No. 7.5;  Indels 1;  Gaps 1;
Matches 114;  Conservative 0;  Mismatches 117;  Indels 1;  Gaps 1;

Qy   213 CCAGGGGGACCTGCTACTTGCCCCCGAGGCTTCGCCGTCACCGGTCACTTGTGGCT 272
         ||  |||||  ||| |  |||   ||| |||  |||||  ||||||  |  | |
Db 53470 CCGGCGNGGCGCCCCGGCGGGGGGGGGGGGCGGNGGCCGGNGGCCGGGGGCCGGGGNC 53411

Qy   273 CCGCCTGTGGCTCGTGGGATGTCGCGCCGAGACCACATGTCACTGCCAGTGCCC-GGGC 331
         |||  |  | | ||| ||   |   | || |   |    |  || ||  | |||| ||||
Db 53410 GCGCCCCGGCGGGGGGCGGGGGGNGGGGCGGGGCCGGGGNCNNCGGGNCNNCGGGGGGGG 53351

Qy   332 ATGGACTGGACCGAGCCGCCTGCTGCTGTGTAGCCCTGAGGTCGCCGCAGCCGCGTG 391
         ||  | | ||    ||       |||| |  |      |||||  ||| |  | | | ||
Db 53350 GGGGGGCGGGGGNNGCCCGGGNCCCGGGGCGCTCCAGGTCGGAGGGGTTGCGGGGAGCTGG 53291

Qy   392 CACAGCGCCGCGGAGGCCGCTCCAGGTCCGAGGGGGTTGCGGGGGAGCTGG 443
         ||| |||||||||   ||    |||  || ||  ||||  |||| |||| ||
Db 53290 GGGGGCGCCCCGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGCCGG 53239


RESULT 15
AC005445
LOCUS       AC005445      75135 bp    DNA              HTG       10-FEB-2000
DEFINITION  Drosophila melanogaster chromosome 2 clone DS07435 (D359) map
            44A1-44A2 strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 6
            unordered pieces.
ACCESSION   AC005445
VERSION     AC005445.4  GI:6957993
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 75135)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Cieslolka,C.,
            Doyle.C.M., Farfan,D.E., Galle.R., George.R.A., Harris,N.L.,
            Hinkle.A., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Kearney,L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 75135)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Feb 11, 2000 this sequence version replaced gi:5655727.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
```

```
the following cutoffs: length >= 200 bases. P1 library location:
43-78.
     * NOTE: This is a 'working draft' sequence. It currently
     * consists of 6 contigs. The true order of the pieces
     * is not known and their order in this sequence record is
     * arbitrary. Gaps between the contigs are represented as
     * runs of N, but the exact sizes of the gaps are unknown.
     * This record will be updated with the finished sequence
     * as soon as it is available and the accession number will
     * be preserved.
     *         1       633: contig of 633 bp in length
     *       634       713: gap of unknown length
     *       714     41434: contig of 40721 bp in length
     *     41435     41514: gap of unknown length
     *     41515     73291: contig of 31777 bp in length
     *     73292     73371: gap of unknown length
     *     73372     73646: contig of 275 bp in length
     *     73647     73726: gap of unknown length
     *     73727     74390: contig of 664 bp in length
     *     74391     74470: gap of unknown length
     *     74471     75135: contig of 665 bp in length.
FEATURES             Location/Qualifiers
     source          1..75135
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2"
                     /map="44A1-44A2"
                     /clone="DS07435 (D359)"
                     /clone_lib="P1 library, partial Sau3A in pAd10sacBII"
BASE COUNT    19834 a  17070 c  17499 g  20325 t   407 others
ORIGIN

Query Match          9.1%;   Score 41.2;  DB 55;  Length 75135;
Best Local Similarity  47.8%;   Pred. No. 7.4;
Matches 109;  Conservative   0;  Mismatches 119;  Indels   0;  Gaps   0;

Qy    216 GGGGGGACCTGGCTACTTGCCCCCGAGGCTTCGCCGTCACCGGCTGCACTTGTGGCTCCG 275
          ||||||  ||  |  ||||||||| |||  |||      ||    |   |||||||| |
Db  73372 GGGGGGGGGGGGGGCCCCCCCCCGGCCGCGCGGGGGGGCCGGCGGGGGGGGGGCCCGG 73431

Qy    276 CCTGTGGCTCGTGGGATGTGCGCCGAGACCACATGTCACTGCCAGTGCGCGGGCATGG 335
          |||||  | ||  |   |||||  |  || |  | | || |||   | |||||||||
Db  73432 CGCCGGGNGGCGGGGCCCCCGCGGGGCCCGGCGGGGGAGCGCGNGNGCGGGGGGGCG 73491

Qy    336 ACTGGACCGGAGCGGCGCTGCTGTCGTGTGTGCAGCCCTGAGGTGCGCGCAGCCGTGCACA 395
          ||  ||||||  |||  |  |       |||  ||| ||   | | ||| | |  || ||
Db  73492 GGGGGCCCGGGGGGCCCCGCGAAAGGGGGCGCCGCGGCGGCCGGGGGCCGGG 73551

Qy    396 GCGCGGGCCGGAGGCCGCTCCAGGTCCGGAGGGGTTGCGGGGGAGCTGG 443
          | ||||||  ||  ||  | | ||||| ||     || ||||  | ||
Db  73552 GGCCGGGGGGGGGGGGGGGGGGCCGGNGGCGGGGGGGGGGGCCCCGCGCGGNGG 73599
```

Search completed: May 26, 2000, 00:32:46
Job time: 2741 sec

/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I - Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCGAAGTGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT   107 a   134 c   122 g   121 t
ORIGIN

Query Match   33.3%;   Score 150.8; DB 38;   Length 484;
Best Local Similarity 67.5%;   Pred. No. 1.6e-25;
Matches 212; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

```
Qy   54 CTCTCGTCTCCTCCCCTCCCCTGTCCCGGGCGTGTTGGTGTCTAGCAAGACCCTGTGCT 113
           |||| | ||| | ||||||| | ||||| ||| |  |||| |  |||  ||| | |||||
Db   90 CCCTCCTTTTCCTTTTCTTCTTGTCCCTGAACTGCTGGGCTCCAGCATGCCACTGTGTC 149

Qy  114 CCATGGAAGAAGCCATCAATGAGAGGATCCAGGAGGTCGCCGGCTCCTAATATTTAGGG 173
           |||| | ||||||||| |   |||  ||||  | ||  |  ||  |  |||  |||||
Db  150 CCATCGATGAAGCCATCGACACAAGAAGATCAAACAAGACTTCAACTCCCTGTTTCCAAATG 209

Qy  174 CAATAAGCAGCATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGGACCTGGCTACTT 233
           ||||||  | ||||||| | || ||| ||| | || ||| ||||  |  |||  |  ||
Db  210 CAATAAAGAACATTGGCTTAAATTGCTGGACAGTCTCCTCCAGAGGGAAGTTGGCCTCCT 269

Qy  234 GCCCCCGAGGCTTCGCCGTCACCGGCTCCGTCCGTCTGGCTCGTGGCTCGTGGGGATG 293
           |||  | |||  | || | || ||||||| | |||||||||| |||||||| |||||
Db  270 GCCCAGAAGGCACAGCAGTCTTGAGCTGCTGCTCCGTCTGGCTCGTGGCTCGTGGGACA 329

Qy  294 TGGCGCCGAGACCACCATGTCACTGCCAGTGCGCGGGCATGGCCGGGAGCCGCGCT 353
           | ||  || ||| | ||||||||||||||||| ||| ||| | | |  | |||||
Db  330 TTCGTGAAGAAAAGTGTGTCACTGCCAGTGTGCAAGGATAGACTGGACAGCAGCGCCCGCT 389

Qy  354 GCTGTCGTGTGCAG 367
           ||||| | |||||
Db  390 GCTGTAAGCTGCAG 403
```

RESULT 7
AA823170
LOCUS   AA823170   508 bp   mRNA   EST   17-FEB-1998
DEFINITION vw41e10.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone IMAGE:1246410 5', mRNA sequence.
ACCESSION   AA823170
VERSION   AA823170.1 GI:2893038
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 508)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Sep 12, 1996 this sequence version replaced gi:1394966.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

BASE COUNT   122 a   157 c   126 g   129 t   1 others
ORIGIN   Bonaldo."

Query Match   33.6%;   Score 152.4; DB 47;   Length 535;
Best Local Similarity 67.8%;   Pred. No. 7e-26;
Matches 213; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

```
Qy   54 CTCTCGTCTCCTCCCCTCCCCTGTCCCGGGCGTGTTGGTGTCTAGCAAGACCCTGTGCT 113
           |||| | ||| | ||||||| | ||||| ||| |  |||| |  |||  ||| | |||||
Db   81 CCCTCCTTTTCCTTTTCTTCTTGTCCCTGAACTGCTGGGCTCCAGCATGCCACTGTGTC 140

Qy  114 CCATGGAAGAAGCCATCAATGAGAGGATCCAGGAGGTCGCCGGCTCCTAATATTTAGGG 173
           |||| | ||||||||| |   |||  ||||  | ||  |  ||  |  |||  |||||
Db  141 CCATCGATGAAGCCATCGACACAAGAAGATCAAACAAGACTTCAACTCCCTGTTTCCAAATG 200

Qy  174 CAATAAGCAGCATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGGACCTGGCTACTT 233
           ||||||  | ||||||| | || ||| ||| | || ||| ||||  |  |||  |  ||
Db  201 CAATAAAGAACATTGGCTTAAATTGCTGGACAGTCTCCTCCAGAGGGAAGTTGGCCTCCT 260

Qy  234 GCCCCCGAGGCTTCGCCGTCACCGGCTCCGTCCGTCTGGCTCGTGGCTCGTGGGGATG 293
           |||  | |||  | || | || ||||||| | |||||||||| |||||||| |||||
Db  261 GCCCAGAAGGCACAGCAGTCTTGAGCTGCTGCTCCGTCTGGCTCGTGGCTCGTGGGACA 320

Qy  294 TGGCGCCGAGACCACCATGTCACTGCCAGTGCGCGGGCATGGCCGGGAGCCGCGCT 353
           | ||  || ||| | ||||||||||||||||| ||| ||| | | |  | |||||
Db  321 TTCGTGAAGAAAAGTGTGTCACTGCCAGTGTGCAAGGATAGACTGGACAGCAGCGCCCGCT 380

Qy  354 GCTGTCGTGTGCAG 367
           ||||| | |||||
Db  381 GCTGTAAGCTGCAG 394
```

RESULT 6
AA796118
LOCUS   AA796118   484 bp   mRNA   EST   09-FEB-1998
DEFINITION vo66e08.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone IMAGE:1054886 5', mRNA sequence.
ACCESSION   AA796118
VERSION   AA796118.1 GI:2855073
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 484)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Jan 17, 1998 this sequence version replaced gi:2044301.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:586462
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES   Location/Qualifiers
     source   1..484
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1054886"
                     /clone_lib="Soares_mammary_gland_NbMMG"

```
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:660098
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 454.

FEATURES           Location/Qualifiers
     source        1..508
                   /organism="Mus musculus"
                   /strain="C57BL/6J"
                   /db_xref="taxon:10090"
                   /clone="IMAGE:1246410"
                   /clone_lib="Soares_mammary_gland_NbMMG"
                   /sex="male"
                   /tissue_type="mammary gland"
                   /dev_stage="4 weeks"
                   /lab_host="DH10B"
                   /note="Organ: mammary gland; Vector: pT7T3D-Pac
                   (Pharmacia) with a modified polylinker; Site_1: Not I;
                   Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                   oligo(dT) primer [5'
                   TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTTT
                   T 3']; double-stranded cDNA was ligated to Eco RI
                   adaptors (Pharmacia), digested with Not I and cloned into
                   the Not I and Eco RI sites of  the modified pT7T3 vector.
                   RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                   constructed and normalized by Bento Soares and M.Fatima
                   Bonaldo."

BASE COUNT      113 a      149 c      122 g      124 t
ORIGIN

Query Match           33.3%;       Score 150.8;  DB 39;  Length 508;
Best Local Similarity 67.5%;       Pred. No. 1.6e-25;
Matches   212;      Conservative    0;  Mismatches  102;  Indels   0;  Gaps   0;

Qy   54 CTCTCTGTCTCCTCCTCCCCCCCGTGTCCTGGGGCGTGTTGGTGTCTAGCAAGACCCTGTGCT 113
        | ||| |||||||  |||| |||||  | ||| |||  |||| || | ||||| |||||
Db   58 CCCTCCTTTCTTTCTTCCTTGTCCCTGAACTGCTGGGCTCCAGCATGCCACTGTGTC 117

Qy  114 CCATGGAAGAAGCCATCAATGAGGATCCAGGAGGTCGCCGGCTCCCTAATATTTAGGG 173
        ||| ||| || |||||||  | ||| | | || |  | ||| || | |  | | |
Db  118 CCATCGATGAAGCCATCGACACGAGAGATCAAACAAGACTTCAACTCCCTGTTTCCAAATG 177

Qy  174 CAATAAGCAGCAGATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTACTT 233
        ||||||  |||||||  | | | ||||  | ||  || |||  ||||||| || | |||
Db  178 CAATAAGAACATTGGCTTAAATTGCTGGACAGCTCTCCTCCAGAGGGAAGTTGGCCTCCT 237

Qy  234 GCCCCCGAGGCTTCGCCGTCACCGGCTGCACTTGTGGCTCCGCCTGTGGCTCGTGGGATG 293
        |||| |||| ||| ||  || ||| |||  | | ||||| |  ||| | |||||||||||
Db  238 GCCCAGAAGGCACAGCAGTCTTGAGCTGCTCCATGTGCCTCGTGGCTCGTCGTGGGACA 297

Qy  294 TGCGGCGCGAGACCACATGTCACTGCCAGTGCCGGGCATGGACCGGAGCGCGCCGCT 353
        | ||| ||| |  | ||||  |||  ||| | ||| | ||| |  ||| | | |||
Db  298 TTCGTGAAGAAAAAGTGTGTCACTGCCAGTGCCGGGCTGCCAGTGACTGGACAGCAGCCCGCT 357

Qy  354 GCTGTCTGCTGCAG 367
        ||||| |||||
Db  358 GCTGTAAGCTGCAG 371

RESULT 8
AW319660/c
LOCUS       AW319660       383 bp    mRNA         EST       25-JAN-2000
DEFINITION  un01e05.x1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2352992 3', mRNA sequence.
ACCESSION   AW319660
VERSION     AW319660.1  GI:6749204
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 383)
AUTHORS     Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
TITLE       The WashU-NCI Mouse EST Project 1999
JOURNAL     Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3137857.
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1012228
            Seq primer: custom primer used
            High quality sequence stop: 335.

FEATURES          Location/Qualifiers
     source       1..383
                  /organism="Mus musculus"
                  /strain="C57BL"
                  /db_xref="taxon:10090"
                  /clone="IMAGE:2352992"
                  /clone_lib="Sugano mouse kidney mkia"
                  /sex="female"
                  /dev_stage="adult"
                  /lab_host="DH10B"
                  /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                  (CACTGTG); Site_2: DraIII (CACCATG); 1st strand cDNA
                  was primed with an oligo(dT) primer
                  [ATGTGGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                  ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested
                  and cloned into distinct DraIII sites of the pME18S-FL3
                  vector (5' site CACCATGTG, 3' site CACCATGTG). XhoI should
                  be used to isolate the cDNA insert.  Size selection was
                  performed to exclude fragments <1.5kb. Library
                  constructed by Dr. Sumio Sugano (University of Tokyo
                  Institute of Medical Science).  Custom primers for
                  sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                  primer CGACCTGCAGCTCGAGCACA."

BASE COUNT      97 a      89 c     108 g      89 t
ORIGIN

Query Match           33.1%;       Score 149.8;  DB 80;  Length 383;
Best Local Similarity 68.2%;       Pred. No. 2.5e-25;
Matches   208;      Conservative    0;  Mismatches  97;  Indels   0;  Gaps   0;

Qy   54 CTCTCTGTCTCCTCCTCCCCCCCGTGTCCTGGGGCGTGTTGGTGTCTAGCAAGACCCTGTGCT 113
        | ||| |||||||  |||| |||||  | ||| |||  |||| || | ||||| |||||
Db  305 CCCTCCTTTTCTTTCTTCCTTGTCCCTGAACTGCTGGGCTCCAGCATGCCACTGTGTC 246

Qy  114 CCATGGAAGAAGCCATCAATGAGGATCCAGGAGGTCGCCGGCTCCCTAATATTTAGGG 173
        ||| ||| || |||||||  | ||| | | || |  | ||| || | |  | | |
Db  245 CCATCGATGAAGCCATCGACACGAGAGATCAAACAAGACTTCAACTCCCTGTTTCCAAATG 186

Qy  174 CAATAAGCAGCAGATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTACTT 233
        ||||||  |||||||  | | | ||||  | ||  || |||  ||||||| || | |||
Db  185 CAATAAGAACATTGGCTTAAATTGCTGGACAGCTCTCCTCCAGAGGGAAGTTGGCCTCCT 126

Qy  234 GCCCCCGAGGCTTCGCCGTCACCGGCTGCACTTGTGGCTCCGCCTGTGGCTCGTGGGATG 293
        |||| |||| ||| ||  || ||| |||  | | ||||| |  ||| | |||||||||||
Db  125 GCCCAGAAGGCACAGCAGTCTTGAGCTGCTCCATGTGCCTCGTGGCTCGTCGTGGGACA 66

Qy  294 TGCGGCGCGAGACCACATGTCACTGCCAGTGCCGGGCATGGACCGGAGCGCGCCGCT 353
        | ||| ||| |  | ||||  |||  ||| | ||| | ||| |  ||| | | |||
Db   65 TTCGTGAAGAAAAAGTGTGTCACTGCCAGTGCCGGGCTGCCAGTGACTGGACAGCAGCCCGCT 6
```

```
Qy  114 CCATGGAAGCCATCAATGAGAGGATCCAGGAGGTCGCCGGCCTCCCTAATATTAGG 173
        |||| || |||||||||||  | ||| |||||||||    |||   | |||
Db  141 CCATCGATGAAGCCATCGACAAGAAGATCAAACAAGACTTCAACTCCCTGTTCCAAATG 200

Qy  174 CAATAAGACAGCATGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTACTT 233
        |||||| || || ||||| | ||| |||| ||| |||  |||||  | | ||||  |
Db  201 CAATAAAGAACATTGGCTTAAATTGCTGGACAGTCTCCTCCAGAGGAGAAGTTGGCCTCCT 260

Qy  234 GCCCCGGAGGCTTCGCCGTACCGGCTGCACTTTGGCGTCCGGCTCCTGTGGCTCGTCGGGATG 293
        |||| |||||| | || |||| ||| |||    |||| |||||||| || ||||||| ||||
Db  261 GCCCAGAAGGCACAGCAGTCTTGAGCTGCCTGCCTGGCTGCCTGTGGCTCGTCGGGACA 320

Qy  294 TGGGCGCGGAGACCACCATGTCACTGCCAGTGCGCGGGGCATGGGACTGGACCGGAGCCGGCGCT 353
        | || | |  | |||||| |||||||| ||||| ||| | || |||||| | ||||| |||  |||
Db  321 TTCGTGAAGAAAAGTGTGTCACTGCCCAGTGTGCAAGGATAGACTGGACAGCAGCCCCGCT 380

Qy  354 GCTGT 358
        |||||
Db  381 GCTGT 385
```

RESULT 9
AI746650
```
LOCUS       AI746650      386 bp    mRNA              EST       22-JUN-1999
DEFINITION  uo1g08.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2064830 5', mRNA sequence.
ACCESSION   AI746650
VERSION     AI746650.1  GI:5124914
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 386)
AUTHORS     Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
TITLE       The WashU-NCI Mouse EST Project 1999
JOURNAL     Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138295.
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            WashU-NCI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:994018
            Putative full length read
            vector to vector length is
            Seq primer: custom primer used.
FEATURES    Location/Qualifiers
     source 1..386
            /organism="Mus musculus"
            /strain="C57BL"
            /db_xref="taxon:10090"
            /clone="IMAGE:2064830"
            /clone_lib="Sugano mouse kidney mkia"
            /sex="female"
            /dev_stage="adult"
            /lab_host="DH10B"
            /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
            (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
            was primed with an oligo(dT) primer
            [ATGTGGCCTTTTTTTTTTTTTTTTTTT]; double-stranded cDNA was
            ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested
            and cloned into distinct DraIII sites of the pME18S-FL3
            vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
            be used to isolate the cDNA insert. Size selection was
            performed to exclude fragments <1.5kb. Library
            constructed by Dr. Sumio Sugano (University of Tokyo
            Institute of Medical Science). Custom primers for
            sequencing; 5' end primer CTTCGCTCTAAAAGCTGCG and 3' end
            primer CGACCTGCAGCTCGAGCACA."
BASE COUNT  90 a   107 c   91 g   98 t
ORIGIN
```

```
Query Match          33.1%;   Score 149.8;   DB 51;   Length 386;
Best Local Similarity 68.2%;   Pred. No. 2.5e-25;
Matches   208;   Conservative 0;   Mismatches 97;   Indels 0;   Gaps 0;

Qy  54 CTCTCTGCTCCTCCTCCCCTGTCCTGGGCTGTTGGTGTCTGCTAGCCAAGACCCTGTGCT 113
       |||| |||| |||||| ||| ||| |||||||| |||| ||| || |||||| |||||||
Db  81 CCCTCCTTTCCTTTTTCTTTTCCTTGTCCCCTGAACTGCGTGGGGTCCCAGCATGCCACTGTGTC 140
```

RESULT 10
AI121371
```
LOCUS       AI121371      432 bp    mRNA              EST       02-SEP-1998
DEFINITION  uc30h11.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:1399557 5', mRNA sequence.
ACCESSION   AI121371
VERSION     AI121371.1  GI:3521695
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 432)
AUTHORS     Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE       The WashU-HHMI Mouse EST Project
JOURNAL     Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1400867.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:911273
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 426.
FEATURES    Location/Qualifiers
     source 1..432
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="IMAGE:1399557"
            /clone_lib="Soares_mammary_gland_NbMMG"
            /sex="male"
            /tissue_type="mammary gland"
            /dev_stage="4 weeks"
            /lab_host="DH10B"
            /note="Organ: mammary gland; Vector: pT7T3D-Pac
            (Pharmacia) with a modified polylinker; Site_1: Not I;
            Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
            oligo(dT) primer [5'-
            TGTTACCAATCGAAGTGGAGGAGGCGCCGAATGTTTTTTTTTTTTTTTTTT
            T 3']; double-stranded cDNA was ligated to Eco RI
            adaptors (Pharmacia), digested with Not I and cloned into
            the Not I and Eco RI sites of the modified pT7T3 vector.
```

RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

BASE COUNT   99 a   117 c   110 g   106 t
ORIGIN

Query Match   31.7%;   Score 143.4;   DB 42;   Length 432;
Best Local Similarity   66.9%;   Pred. No. 7.9e-24;
Matches   204; Conservative   0; Mismatches 101; Indels   0; Gaps   0;

Qy   54 CTCTCGTCTCCTCCCTCCCTGTCCTGGGCTGTTGGTGTCTAGCAAGACCCTGTGCT 113
              | ||| |||| ||| |||| | ||||| | |||||| |||| |||||| ||||||
Db   85 CCCTCCTTTCCTTTCTTCCTTGTCCCTGAACTGCTGGCTCCAGCATGCCACTGTGTC 144

Qy   114 CCATGGAAGAAGCCATCAATGAGAGGATCCAGGAGGTCCCGGCTCCCTAATATTTAGGG 173
              |||| || ||||||||| | ||| |||| | | | | | |||||| | |
Db   145 CCATCGATGAAGCCATCGACACAGAAGATCAAACAAGACTTCAACTCCCTGTTTCCAAATG 204

Qy   174 CAATAAGCAGCATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTACTT 233
              ||||| | || || ||||| |||| || || |||||| || ||||||| ||| |
Db   205 CAATAAAGAACATTGGCTTAAATTGCTGGACAGTCTCCTCCAGAGGGAAGTTGGCCTCCT 264

Qy   234 GCCCCCGAGGCTTCGCGTCACCGGCTGCACTTGTGCTCCGCCTGTGGCTCCGCTGGGGATG 293
              ||| ||| |||| ||| || | ||||||| ||||||||||||||||||||||||||
Db   265 GCCCAGAAGGCACAGCAGTCTTGAGCTGCTCCTGTGGGCTCTG-CTGTGGCTCGTGGGACA 324

Qy   294 TGCGGCCGAGACCACCATGTCACTGCCAGTGCCGGGCATGGACTGGACGGACGGCGGCT 353
              ||| | | | |||| | ||||||||||| | |||||||||||||||| | ||||||
Db   325 TTCGTGAAGAAAAGTGTCACTGCCAGTGTCACTGCCAGTGTGCAAGGATAGACTGGACAGCAGCCCGCT 384

Qy   354 GCTGT 358
              |||
Db   385 CGTGT 389

RESULT 11
AA139082
LOCUS   AA139082   493 bp   mRNA   EST   16-FEB-1997
DEFINITION   mr11g09.r1 Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:597184
             5', mRNA sequence.
ACCESSION   AA139082
VERSION   AA139082.1 GI:1701283
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 493)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Sep 12, 1996 this sequence version replaced gi:1397708.
             Contact: Marra M/Mouse EST Project
             WashU-HHMI Mouse EST Project
             Washington University School of MedicineP
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:362616
             Seq primer: -28M13 rev2 from Amersham
             High quality sequence stop: 448.
FEATURES   Location/Qualifiers
     source 1..493
             /organism="Mus musculus"
             /strain="C57BL/6J"

/db_xref="taxon:10090"
/clone="IMAGE:597184"
/clone_lib="Soares mouse 3NbMS"
/sex="male"
/tissue_type="spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCGAAGTGGGAGCGGCCGCCGTGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Fatima Bonaldo."

BASE COUNT   110 a   139 c   122 g   122 t
ORIGIN

Query Match   31.0%;   Score 140.4;   DB 29;   Length 493;
Best Local Similarity   67.5%;   Pred. No. 4.1e-23;
Matches   212; Conservative   0; Mismatches 101; Indels   1; Gaps   1;

Qy   54 CTCTCGTCTCCTCCCTCCCTGTCCTGGGCTGTTGGTGTCTAGCAAGACCCTGTGCT 113
              | ||| |||| ||| |||| | ||||| | |||||| |||| |||||| ||||||
Db   83 CCCTCCTTTCCTTTCTTCCTTGTCCCTGAACTGCTGGCTCCAGCATGCCACTGTGTC 142

Qy   114 CCATGGAAGAAGCCATCAATGAGAGGATCCAGGAGGTCCCGGCTCCCTAATATTTAGGG 173
              |||| || ||||||||| | ||| |||| | | | | | |||||| | |
Db   143 CCATCGATGAAGCCATCGACACAGAAGATCAAACAAGACTTCAACTCCCTGTTTCCAAATG 202

Qy   174 CAATAAGCAGCATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTACTT 233
              ||||| | || || ||||| |||| || || |||||| || ||||||| ||| |
Db   203 CAATAAAGAACATTGGCTTAAATTGCTGGACAGTCTCCTCCAGAGGGAAGTTGGCCTCCT 262

Qy   234 GCCCCCGAGGCTTCGCGTCACCGGCTGCACTTGTGCTCCGCCTGTGGCTCCGCTGGGGATG 293
              ||| ||| |||| ||| || | ||||||| ||||||||||||||||||||||||||
Db   263 GCCCAGAAGGCACAGCAGTCTTGAGCTGCTCCTCTG-CTGTGGCTCGTGG-CTGTGGGGACA 321

Qy   294 TGCGGCCGAGACCACCATGTCACTGCCAGTGCCGGGCATGGACTGGACGGACGGCGGCT 353
              ||| | | | |||| | ||||||||||| | |||||||||||||||| | ||||||
Db   322 TTCGTGAAGAAAAGTGTCACTGCCAGTGTCACTGCCAGTGTGCAAGGATAGACTGGACAGCAGCCCGCT 381

Qy   354 GCTGTCGTGTGCCAG 367
              ||||| |||||
Db   382 GCTGTAAGCTGCAG 395

RESULT 12
W42057
LOCUS   W42057   379 bp   mRNA   EST   11-SEP-1996
DEFINITION   mb16d01.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
             IMAGE:329569 5', mRNA sequence.
ACCESSION   W42057
VERSION   W42057.1 GI:1326540
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 379)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Apr 14, 1993 this sequence version replaced gi:716839.
             Contact: Marra M/Mouse EST Project
             WashU-HHMI Mouse EST Project

Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:210969

Seq primer: ETprimer
High quality sequence stop: 373.

FEATURES
source   Location/Qualifiers
                1..379
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="IMAGE:329569"
                /clone_lib="Soares mouse p3NMF19.5"
                /dev_stage="19.5 dpc total fetus"
                /lab_host="DH10B (ampicillin resistant)"
                /note="Vector: pT7T3D (Pharmacia) with a modified
                polylinker; Site_1: Not I - oligo(dT) primer [5'
                TGTTACCAATCTGAAGTGGGAGGCGGCCGCATTTTTTTTTTTTTTTTTTT 3'],
                double-stranded cDNA was size selected, ligated to Eco RI
                adapters (Pharmacia), digested with Not I and cloned into
                the Not I and Eco RI sites of a modified pT7T3 vector
                (Pharmacia). Library went through one round of
                normalization to a Cot = 5. Library constructed by Bento
                Soares and M.Fatima Bonaldo. RNA was kindly provided by
                Dr. Minoru Ko (Wayne State University)."

BASE COUNT   91 a   104 c   89 g   95 t
ORIGIN

Query Match   30.1%;   Score 136.2;   DB 26;   Length 379;
Best Local Similarity 66.6%;   Pred. No. 3.5e-22;
Matches 195;   Conservative   0;   Mismatches 98;   Indels   0;   Gaps   0;

Qy   54 CTCTCTGTCTCCCTCTCCCCTCTGGGCTGTGGTGTCTAGCAAGACCCTGCT 113
          |||||||| ||||| ||||||||||||||||||||||| ||||||||||||
Db   87 CCCTCCTTTTCCTTTTCTTTCTTGTCCCTGAACTGCTGGGCTCCAGCATGCCACTGTGTC 146

Qy   114 CCATGGAAGAAGCCATCAATGAGAGGATCCAGGAGGTCGCCGGCTCCCTAATATTTAGG 173
          ||| ||||||||||||||| ||| ||||| |||| ||| || |||| |
Db   147 CCATCGATGAAGCCATCGACAAGAAGATCAAACAAGACTTCAACTCCCTTTTCCAAATG 206

Qy   174 CAATAAGCAGCATTGGCCTGAGTGCCCAGAGCGTCACCTCCAGGGGGGACCTGGCTACTT 233
          |||||| |||||||||| || || |||| |||||| || ||||||||||||||
Db   207 CAATAAAGAACATTGGCTTAAATTGCTGGACAGTCTCCTCCAGAGGGGAAGTGGCCTCCT 266

Qy   234 GCCCCCGAGGCTTCGGCCTCACCGGCTGCACTTGTCCTGCCGCCTCGTGGGCTCGTGGGATG 293
          |||| |||||||| ||| ||| ||||||||||| |||||||
Db   267 GCCCAGAAGGCACAGCAGTCTTAGACTGTGACAGTCTCCTCCAGAGGGGAAGTGGCCTCCT 326

Qy   294 TGCGCGCCGAGACCACCACATGTCACTGCCAGTGCGGGGCATGGACTGGACCGGA 346
          | || ||||| || || ||||||||||||||||| || |||||||
Db   327 TTCGTGAAGAAAAAAGTGTCACTGCCAGTGCGAAGGATAGACTGGACAGCAGGCA 379

RESULT 13
AI021019
LOCUS   AI021019   496 bp   mRNA   EST   16-JUN-1998
DEFINITION ua99a04.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:1365582 5', mRNA sequence.
ACCESSION   AI021019
VERSION   AI021019.1 GI:3235355
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE   1 (bases 1 to 496)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 458.

FEATURES
source   Location/Qualifiers
                1..496
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="IMAGE:1365582"
                /clone_lib="Soares_mammary_gland_NbMMG"
                /sex="male"
                /tissue_type="mammary gland"
                /dev_stage="4 weeks"
                /lab_host="DH10B"
                /note="Organ: mammary gland; Vector: pT7T3D-Pac
                (Pharmacia) with a modified polylinker; Site_1: Not I; Not I -
                Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                oligo(dT) primer [5'
                TGTTACCAATCTGAAGTGGGAGGCGGCCGAATGTTTTTTTTTTTTTTTTTTT
                T 3']; double-stranded cDNA was ligated to Eco RI
                adapters (Pharmacia), digested with Not I and cloned into
                the Not I and Eco RI sites of the modified pT7T3 vector.
                RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                constructed and normalized by Bento Soares and M.Fatima
                Bonaldo."

BASE COUNT   109 a   140 c   124 g   123 t
ORIGIN

Query Match   29.6%;   Score 134;   DB 41;   Length 496;
Best Local Similarity 67.0%;   Pred. No. 1.2e-21;
Matches 205;   Conservative   0;   Mismatches 100;   Indels   1;   Gaps   1;

Qy   54 CTCTCTGTCTCCCTCTCCCCTCTGGGCTGTGGTGTCTAGCAAGACCCTGTGCT 113
          |||||||| ||||| |||||||||||||||||||||| ||||||||||||
Db   87 CCCTCCTTTTCCTTTTCTTTCTTGTCCCTGAACTGCTGGGCTCCA-CATGCCACTGTGTC 145

Qy   114 CCATGGAAGAAGCCATCAATGAGAGGATCCAGGAGGTCGCCGGCTCCCTAATATTTAGG 173
          ||| ||||||||||||| ||| ||||| |||| ||| || |||| |
Db   146 CCATCGATGAAGCCATCGACAAGAAGATCAAACAAGACTTCAACTCCCTGTTTCCAAATG 205

Qy   174 CAATAAGCAGCATTGGCCTGAGTGCCCAGAGCGTCACCTCCAGGGGGGACCTGGCTACTT 233
          |||||| ||||||||| || || |||| |||||| || ||||||||||||||
Db   206 CAATAAAGAACATTGGCTTAAATTGCTGGACAGTCTCCTCCAGAGGGGAAGTGGCCTCCT 265

Qy   234 GCCCCCGAGGCTTCGGCCTCACCGGCTGCACTTGTCCTGCCGCCTCGTGGGCTCGTGGGATG 293
          |||| |||||||| ||| ||| ||||||||||| |||||||
Db   266 GCCCAGAAGGCACAGCAGTCTTGAGCTGCTCCTGTCCTGCTGCTCGTGGCTCGTGGGACA 325

Qy   294 TGCGCGCCGAGACCACCACATGTCACTGCCAGTGCGGGGCATGGACTGGACCGGAGCCGCT 353
          | || ||||| || || ||||||||||||||| || |||||||| ||||||||
Db   326 TTCGTGAAGAAAAAAGTGTCACTGCCAGTGCAAGGATAGACTGGACAGCAGCCCGCT 385

Qy   354 GCTGTC 359
          ||||||
Db   386 GCTGTC 391

RESULT 14
W42069
LOCUS   W42069   372 bp   mRNA   EST   11-SEP-1996
DEFINITION   mb16d04.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
               IMAGE:329575 5', mRNA sequence.
ACCESSION   W42069
VERSION   W42069.1 GI:1326543
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 372)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
               Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
               Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
               Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
               Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Dec 12, 1995 this sequence version replaced gi:1118196.
               Contact: Marra M/Mouse EST Project
               WashU-HHMI Mouse EST Project
               Washington University School of Medicine
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: mouseest@watson.wustl.edu
               This clone is available royalty-free through LLNL ; contact the
               IMAGE Consortium (info@image.llnl.gov) for further information.
               MGI:210975
               Seq primer: ETprimer
               High quality sequence stop: 364.
FEATURES   Location/Qualifiers
     source   1..372
               /organism="Mus musculus"
               /db_xref="taxon:10090"
               /clone="IMAGE:329575"
               /clone_lib="Soares mouse p3NMF19.5"
               /dev_stage="19.5 dpc total fetus"
               /lab_host="DH10B (ampicillin resistant)"
               /note="Vector: pT7T3D (Pharmacia) with a modified
               polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
               was primed with a Not I - oligo(dT) primer [5'
               TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
               double-stranded cDNA was size selected, ligated to Eco RI
               adapters (Pharmacia), digested with Not I and cloned into
               the Not I and Eco RI sites of a modified pT7T3 vector
               (Pharmacia). Library went through one round of
               normalization to a Cot = 5. Library constructed by Bento
               Soares and M.Fatima Bonaldo. RNA was kindly provided by
               Dr. Minoru Ko (Wayne State University)."

BASE COUNT   90 a   101 c   88 g   93 t
ORIGIN

Query Match   29.3%; Score 132.8; DB 26; Length 372;
Best Local Similarity 66.3%; Pred. No. 2.2e-21;
Matches 191; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy   54 CTCTCTGTCTCCTCCTCCTCCTCCTTCCTTCCTCTGGGCGTGTTGGTGTCTAGCAAGACCCTGTGCT 113
          ||||||| ||||| | |||| || |||||||| |||||| |||||| |||| ||||||||||||| |
Db   85 CCCTCCTTTCCTTTTCCTTTCCTTGTCCCTGTCCCTGAACTGCTGGGCTCCAGCATGCCACTGTGTC 144

Qy   114 CCATGGAGGAAGGCCATCAATGAGAGGATCCAGGAGGTCGCCGGCTCCCTAATATTTAGGG 173
          |||| || ||||| |||||||||||| |||| ||||| | ||||||| | ||||||||
Db   145 CCATCGATGAAGGCCATCGACAAGAGAGATCAAACAGAGACTTCAACTCCCCGTTTCCAAATG 204

Qy   174 CAATAAGCAGCATTGGCCTGGAGTGCCAGAGCCGTCACCTCCAGGGGGGGACCTGGCTACTT 233
          ||||| |||| |||| |||||| ||||||| ||| || |||||||||| ||| ||||| |
Db   205 CAATAAAGAACATTGGCTAAATTGCTGGACAGTCTCCTCCCAGAGGAAGTTGGCCTCCT 264

Qy   234 GCCCCCGAGGCTTCGCCGTCCACCCGGCTCACTTGTGGCTCCGCCTGTTGGTCTCGTGGGATG 293

Db   265 |||| |||| |||| ||||| ||||||| ||||| || |||| |||||| ||| |

Qy   294 TGCGCCGCCGACCACCATGTCACTGCCAGTGCGCGGGGGCATGGCGACTGGA 341
          |||||||||||||||| || ||||||||||||||||||||||||||||||||

Db   325 TTCGTGAAGAAAAAGTGTGTCACTGCCAGTGTCAAGGATAGAACTGGA 372

RESULT 15
AA423511
LOCUS   AA423511   445 bp   mRNA   EST   16-OCT-1997
DEFINITION   ve76b03.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
               IMAGE:832109 5', mRNA sequence.
ACCESSION   AA423511
VERSION   AA423511.1 GI:2102329
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 445)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
               Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
               Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
               Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
               Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On May 5, 1995 this sequence version replaced gi:798456.
               Contact: Marra M/Mouse EST Project
               WashU-HHMI Mouse EST Project
               Washington University School of Medicine
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: mouseest@watson.wustl.edu
               This clone is available royalty-free through LLNL ; contact the
               IMAGE Consortium (info@image.llnl.gov) for further information.
               MGI:492325
               Seq primer: -28m13 rev2 ET from Amersham
               High quality sequence stop: 429.
FEATURES   Location/Qualifiers
     source   1..445
               /organism="Mus musculus"
               /strain="C57BL/6J"
               /db_xref="taxon:10090"
               /clone="IMAGE:832109"
               /clone_lib="Soares_mammary_gland_NbMMG"
               /sex="male"
               /tissue_type="mammary gland"
               /dev_stage="4 weeks"
               /lab_host="DH10B"
               /note="Organ: mammary gland; Vector: pT7T3D-Pac
               (Pharmacia) with a modified polylinker; Site_1: Not I;
               Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
               oligo(dT) primer [5'
               TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGTTTTTTTTTTTTTTTTTT
               T 3'], double-stranded cDNA was ligated to Eco RI
               adaptors (Pharmacia), digested with Not I and cloned into
               the Not I and Eco RI sites of the modified pT7T3 vector.
               RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
               constructed and normalized by Bento Soares and M.Fatima
               Bonaldo."

BASE COUNT   102 a   120 c   113 g   110 t
ORIGIN

Query Match   29.1%; Score 132; DB 33; Length 445;
Best Local Similarity 65.9%; Pred. No. 3.5e-21;
Matches 207; Conservative 0; Mismatches 105; Indels 2; Gaps 1;

Qy   54 CTCTCTGTCTCCTCCTCCTCCTCCTTCCTTCCTCTGGGCGTGTTGGTGTCTAGCAAGACCCTGTGCT 113
          | ||| |||| | ||| | ||||||| |||||| | | |||| || | ||||| |||||||||

```
Db    91 CCCTCCTTTCCTTTCTTCCTTGTCCCTGAACTGCTGGGCTCCAGCATGCCACTGTGTC 150

Qy   114 CCATGGAAGAAGCCATCAATGAGAGGATCCAGGAGGTCGCCGGCTCCCTAATATTTAGGG 173
         |||| || |||||||| | ||| |||| | | | ||||||| || | | |
Db   151 CCATCGATGAAGCCATCGACAAGAAGATCAAACAAGACTTCAACTCCCTGTTTCCAAATG 210

Qy   174 CAATAAGCAGCATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTACTT 233
         |||||| | |||||| | ||||||| | ||| ||| |||||||| ||| | |||| ||
Db   211 CAATAAAGAAACATTGGCTTAAATTGCTGGACAGTCTCCTCCAGAGGGAAGTTGGCCTCCT 270

Qy   234 GCCCCCGAGGCTTCGCCGTCACCGGCTGCACTTGTGGCTCCGCCTGTGGCTCGTGGGATG 293
         |||| |||| |||| ||||| | ||||||| | | ||||||||
Db   271 GCCCAGAAGGCACACAGCAGTCTTGAGCTGCTCCTGTGGCTCTGC--TGTGGTCGTGGGACA 328

Qy   294 TGCGCGCCGAGACCACCATGTCACTGCCAGTGCCGCGGGCATGGACTGGACCGGAGCGCGCT 353
         | || |||  |||||||||| || ||||||||| || |||||||  || ||||
Db   329 TTCGTGAAGAAAAGTGTGTCACTGCCAGTGTGCAAGGATAGACTGGACAGCAGCCCCGCT 388

Qy   354 GCTGTCGTGTGCAG 367
         ||||| |||||
Db   389 GCTGTAAGCTGCAG 402


Search completed: May 26, 2000, 00:15:27
Job time: 3157 sec
```

```
                    GenCore version 4.5
              Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   May 26, 2000, 00:28:51 ; Search time 53.36 Seconds
                                    (without alignments)
                                    118.666 Million cell updates/sec

Title:        US-09-099-898-2
Perfect score: 581
Sequence:  1 MKALCLLLLPVLGLLVSSKT........CHCQCAGMDWTGARCCRVQP 108

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:  168808 seqs, 58629743 residues

Total number of hits satisfying chosen parameters:    168808

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database : PIR_63:*
           1:  pirl:*
           2:  pir2:*
           3:  pir3:*
           4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	86	14.8	934	1	A34372	complement C6 prec
2	86	14.8	1059	2	T22545	hypothetical prote
3	82.5	14.2	1574	2	T13954	MEGF6 protein - ra
4	82	14.1	2321	2	S78549	notch3 protein - h
5	80.5	13.9	456	1	KXBO	protein C (activat
6	78	13.4	1106	2	T18739	hypothetical prote
7	75	13.3	295	2	JC5559	lectin-B - Virgini
8	75.5	13.0	387	2	B49175	Motch A protein -
9	75.5	13.0	2139	2	A35672	crumbs protein - f
10	75.5	13.0	2531	2	A46019	Notch-1 protein -
11	75	12.9	124	2	A21761	high-cysteine chor
12	75	12.9	748	2	T37097	probable secreted
13	74.5	12.8	2555	2	A40043	notch protein homo
14	74	12.7	2531	2	S18188	notch protein homo
15	73.5	12.7	289	2	T25682	hypothetical prote
16	73.5	12.7	660	2	F70101	primosomal protein
17	73.5	12.7	1620	2	T27283	hypothetical prote
18	73	12.6	861	2	A48825	Notch homolog Motc
19	73	12.6	1049	1	S19421	ATP-dependent perm
20	73	12.6	1700	2	S08167	Balbiani ring 3 pr
21	73	12.6	2437	2	S42612	transmembrane prot
22	72	12.4	1111	2	B44018	laminin B2t chain
23	72	12.4	1193	2	A44018	laminin B2t chain
24	72	12.4	2318	2	S45306	notch 3 protein -
25	71.5	12.3	105	2	S23061	chorion protein -
26	71.5	12.3	519	1	YRHUR2	dopachrome Delta-i
27	71.5	12.3	1647	2	T32934	hypothetical prote
28	71.5	12.3	2476	3	T34022	zonadhesin - pig
29	71.5	12.3	2703	1	A24420	notch protein - fr
30	71	12.2	72	2	S39418	metallothionein 10
31	71	12.2	341	2	D48435	cysteine proteinas
32	71	12.2	396	1	TRYXB4	alpha-lytic protei
33	71	12.2	1134	1	JN0711	protein-tyrosine k
34	71	12.2	2907	2	A57278	fibrillin-2 precur
35	70.5	12.1	626	2	T04895	vacuolar sorting r
36	70	12.0	582	2	D72075	hypothetical prote
37	70	12.0	2871	2	A55624	fibrillin-1 precur
38	69.5	12.0	346	2	S34165	keratin, type II -
39	69.5	12.0	384	2	S64735	retrovirus-related
40	69.5	12.0	2524	2	A35844	Xotch protein - Af
41	69	11.9	447	2	A39321	mucin - rat (fragm
42	69	11.9	625	2	T05319	vacuolar sorting r
43	69	11.9	2718	2	A23475	G surface protein
44	68.5	11.8	258	2	T32542	hypothetical prote
45	68.5	11.8	540	2	B47417	insulin receptor-r

ALIGNMENTS

```
RESULT  1
A34372
complement C6 precursor - human
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A34372; A34235; A32109; A31894; A53072
R;Haefliger, J.A.; Tschopp, J.; Vial, N.; Jenne, D.E.
J. Biol. Chem. 264, 18041-18051, 1989
A;Title: Complete primary structure and functional characterization of the sixth comp
A;Reference number: A34372; MUID:90036879
A;Accession: A34372
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-934 <HAE>
A;Cross-references: GB:J05064; NID:g179703; PIDN:AAA51860.1; PID:g179704
R;DiScipio, R.G.; Hugli, T.E.
J. Biol. Chem. 264, 16197-16206, 1989
A;Title: The molecular architecture of human complement component C6.
A;Reference number: A34235; MUID:89380223
A;Accession: A34235
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-118,'E',120-934 <DIS>
A;Cross-references: GB:J05024; NID:g187824; PIDN:AAA59668.1; PID:g307228
R;Chakravarti, D.N.; Chakravarti, B.; Parra, C.A.; Muller-Eberhard, H.J.
Proc. Natl. Acad. Sci. U.S.A. 86, 2799-2803, 1989
A;Title: Structural homology of complement protein C6 with other channel-forming prot
A;Reference number: A32109; MUID:89202413
A;Accession: A32109
A;Molecule type: mRNA
A;Residues: 1-118,'E',120-491 <CH2>
A;Cross-references: GB:J04506; NID:g618465; PIDN:AAB59433.1; PID:g618466
R;Chakravarti, D.N.; Muller-Eberhard, H.J.
J. Biol. Chem. 263, 18306-18312, 1988
A;Title: Biochemical characterization of the human complement protein C6. Association
A;Reference number: A31894; MUID:89054009
A;Accession: A31894
A;Molecule type: protein
A;Residues: 22-41 <CHA>
R;Hobart, M.J.; Fernie, B.; DiScipio, R.G.
Biochemistry 32, 6198-6205, 1993
A;Title: Structure of the human C6 gene.
A;Reference number: A53072; MUID:93291175
A;Accession: A53072
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 'GPGGSQG',1-47,'A',49-118,'E',120-561,'A',563-618,'A',620-700,'A',702-763
A;Note: sequence extracted from NCBI backbone (NCBIP:134071)
C;Genetics:
A;Gene: GDB:C6
A;Cross-references: GDB:119045; OMIM:217050
A;Map position: 5p13-5p13
C;Superfamily: complement c6; agrin inhibitor-like repeat homology; complement factor
```

gy
C;Keywords: complement alternate pathway; complement pathway; cytolysis; glycoprotein;
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-934/Product: complement C6 #status experimental <MAT>
F;22-79/Domain: thrombospondin type 1 repeat homology <THR1>
F;80-134/Domain: thrombospondin type 1 repeat homology <THR2>
F;140-173/Domain: LDL receptor ligand-binding repeat homology <LDL>
F;521-552/Domain: EGF homology <EGF>
F;564-612/Domain: thrombospondin type 1 repeat homology <THR3>
F;644-699/Domain: complement factor H repeat homology <FH01>
F;704-761/Domain: complement factor H repeat homology <FH02>

Query Match   14.8%;   Score 86;   DB 1; Length 934;
Best Local Similarity   29.4%;   Pred. No. 1.4;
Matches   30;   Conservative   10;   Mismatches   36;   Indels   26;   Gaps   7;

Qy   22 CSMEEAINERIQEVAGSLIFRAISSIGLECQSVTSRGDLATCPRGFAVTG---CTCGSAC 78
                      ||:: ||| |: :|| || |||:|||| ||
Db   699 CQRTECIKPVVQEVLTITPFQRLYRIG---ESIE----LTCPKGFVVAGPSRYTCQG-- 749

Qy   79 GSW--DVRAETTC---------HCQCAGMDWTGARCCRVQP 108
             ||| ||||| ||||| |||||| |
Db   749 NSWTPPISNSLTCEKDTLTLKGHCQ-LGQKQSGSECICMSP 789

RESULT 2
T22545
· hypothetical protein F53B6.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T22545
R;White, S.
submitted to the EMBL Data Library, October 1996
A;Reference number: 219578
A;Accession: T22545
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1059 <WIL>
A;Cross-references: EMBL:Z81086; PIDN:CAB03121.1; GSPDB:GN00019; CESP:F53B6.2
A;Experimental source: clone F53B6
C;Genetics:
A;Gene: CESP:F53B6.2
A;Map position: 1
A;Introns: 38/3; 92/3; 131/3; 169/1; 236/2; 282/3; 349/2; 376/2; 420/1; 576/3; 759/3; 84

Query Match   14.8%;   Score 86;   DB 2; Length 1059;
Best Local Similarity   27.2%;   Pred. No. 1.6;
Matches   31;   Conservative   12;   Mismatches   33;   Indels   38;   Gaps   9;

Qy   20 TLCSMEEAINE-------RIQEVAGSLIFRAISSIGLECQSVTSRGDLATCPRGFA-VTG 71
              |::   :|| | :: || : ||| || ||| ||||||||| :
Db   80 TICGGGEIVSRGQCEVVCRSRLTGANFLWRV--DDGTPCQAATSR--AVCSKGSCQIVG 134

Qy   72 C--------TCGSACGS---------WDVRAE-TTCHCQCAG-MDWTGA 101
           | ||||| |||||| ||| ||| |||||
Db   135 CDGLISSSFRFDACG-VCGGRGDTCDNGKFIWKVSEYTACASNCDDIVDWSGA 187

RESULT 3
T13954
MEGF6 protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T13954
R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A;Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A;Reference number: 214126; MUID:98350089
A;Accession: T13954
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA

A;Residues: 1-1574 <NAK>
A;Cross-references: EMBL:AB011532; NID:d1226538; PID:d1033425; PIDN:BAA32462.1
A;Experimental source: strain Sprague-Dawley; brain
C;Genetics:
A;Gene: MEGF6

Query Match   14.2%;   Score 82.5;   DB 2; Length 1574;
Best Local Similarity   35.6%;   Pred. No. 4.6;
Matches   21;   Conservative   3;   Mismatches   26;   Indels   9;   Gaps   2;

Qy   47 IGLECQSVTSRGDLATCPRGFAVTGCTCGSACGSWDVRAETTCHCQCAGMDWTGARCCR 105
            || || | || ||| | ||| ||| |||||||||
Db   813 VGSRCQD-------TCSAGWYGTGCQIRCACANDGHCDPTTGRCSCA-PGWTGLSCQR 862

RESULT 4
S78549
notch3 protein - human
C;Species: Homo sapiens (man)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Sep-1999
C;Accession: S78549; S71825
R;Joutel, A.; Tournier-Lasserve, E.
submitted to the EMBL Data Library, April 1997
A;Reference number: S78549
A;Accession: S78549
A;Molecule type: mRNA
A;Residues: 1-2321 <JOU1>
A;Cross-references: EMBL:U97669; NID:g2668591; PIDN:AAB91371.1; PID:g2668592
R;Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabriat, H.; Mouton, P.; Alamow
x, M.M.; Weissenbach, J.; Bach, J.F.; Bousser, M.G.; Tournier-Lasserve, E.
Nature 383, 707-710, 1996
A;Title: Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stro
A;Reference number: S71825; MUID:97032728
A;Accession: S71825
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 67-113;138-194;268-333;'G',335-346;536-613;716-765;1240-1279;1815-1888.<J
A;Cross-references: EMBL:U97669
C;Genetics:
A;Gene: notch3
A;Map position: 19p13.1
C;Function:
A;Description: may be involved in pathogenesis of CADASIL, causing a type of stroke a
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol
C;Keywords: tandem repeat; transmembrane protein
F;123-155/Domain: EGF homology <EGX1>
F;162-194/Domain: EGF homology <EGF1>
F;240-271/Domain: EGF homology <EGX2>
F;318-349/Domain: EGF homology <EGF2>
F;473-504/Domain: EGF homology <EGX3>
F;853-884/Domain: EGF homology <EGF3>
F;928-959/Domain: EGF homology <EGX4>
F;1070-1126/Domain: laminin-type EGF-like homology <LEG>
F;1838-1870/Domain: ankyrin repeat homology <AN1>
F;1871-1903/Domain: ankyrin repeat homology <AN2>
F;1905-1937/Domain: ankyrin repeat homology <AN3>
F;1938-1970/Domain: ankyrin repeat homology <AN4>
F;1971-2003/Domain: ankyrin repeat homology <AN5>

Query Match   14.1%;   Score 82;   DB 2; Length 2321;
Best Local Similarity   29.9%;   Pred. No. 7;
Matches   23;   Conservative   7;   Mismatches   23;   Indels   24;   Gaps   4;

Qy   47 IGLECQSVTSRGDLATCPRGFAVTGCTCGSACGSWDVRAETTC----------- 90
            :|: || ||||||||||||||||| || :| : ::| :|
Db   1300 VGVPCQQ-TPRGPRCACPPG--LSGPSCRSFPGSPPGASNASCAAAPCLHGGSCRPAPLA 1356

Qy   90 --HCQCAGMDWTGARC 103
             |||||||||||| ||
Db   1357 PFFRCACA-QGWTGPRC 1372

RESULT 5
KXBO

protein C (activated) (EC 3.4.21.69) precursor - bovine (fragment)
N;Alternate names: autoprothrombin IIA; plasma protein C
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Nov-1980 #sequence_revision 17-Mar-1987 #text_change 16-Jul-1999
C;Accession: A26250; A18385; A18386; A00928
R;Long, G.L.; Balagaje, R.M.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 81, 5653-5656, 1984
A;Title: Cloning and sequence of liver cDNA coding for bovine protein C.
A;Reference number: A26250; MUID:85014826
A;Accession: A26250
A;Molecule type: mRNA
A;Residues: 1-456 <LON>
R;Fernlund, P.; Stenflo, J.
J. Biol. Chem. 257, 12170-12179, 1982
A;Title: Amino acid sequence of the light chain of bovine protein C.
A;Reference number: A18385; MUID:83007325
A;Accession: A18385
A;Molecule type: protein
A;Residues: 40-194 <FER>
A;Note: 82-Lys was also found
R;Drakenberg, T.; Fernlund, P.; Roepstorff, P.; Stenflo, J.
Proc. Natl. Acad. Sci. U.S.A. 80, 1802-1806, 1983
A;Title: beta-Hydroxyaspartic acid in vitamin K-dependent protein C.
A;Reference number: A19916; MUID:83169769
A;Contents: annotation; revision to residue 110
R;Stenflo, J.; Fernlund, P.
J. Biol. Chem. 257, 12180-12190, 1982
A;Title: Amino acid sequence of the heavy chain of bovine protein C.
A;Reference number: A18386; MUID:83007326
A;Accession: A18386
A;Molecule type: protein
A;Residues: 197-454, 'PV' <STE>
R;Esmon, N.L.; DeBault, L.E.; Esmon, C.T.
J. Biol. Chem. 258, 5548-5553, 1983
A;Title: Proteolytic formation and properties of gamma-carboxyglutamic acid-domainless p
A;Reference number: A37541; MUID:83213513
A;Contents: annotation; activation; calcium binding
R;Johnson, A.E.; Esmon, N.L.; Laue, T.M.; Esmon, C.T.
J. Biol. Chem. 258, 5554-5560, 1983
A;Title: Structural changes required for activation of protein C are induced by Ca2+ bin
A;Reference number: A37542; MUID:83213514
A;Contents: annotation; activation; calcium binding
C;Comment: Protein C is the zymogen of the vitamin K-dependent serine proteinase that re
s.
C;Comment: Protein C is synthesized in the liver as a single chain precursor, which is c
bin, which cleaves a tetradecapeptide from the amino end of the heavy chain; this reacti
cognition of the thrombin-thrombomodulin complex.
C;Comment: Calcium binds to the gamma-carboxyglutamic acid (Gla) residues and, with stro
C;Comment: The gamma-carboxyglutamic acid residues arise by a posttranslational, vitamin
C;Superfamily: Coagulation factor X; EGF homology; Gla domain homology; trypsin homology
C;Keywords: anticoagulant; beta-hydroxyaspartic acid; blood coagulation; calcium binding
F;1-29/Domain: signal sequence (fragment) #status predicted <SIG>
F;24-83/Domain: Gla domain homology <GLA>
F;30-39/Domain: propeptide #status predicted <PRO>
F;40-194/Product: protein C light chain #status experimental <LCH>
F;98-128/Domain: EGF homology <EG1>
F;137-172/Domain: EGF homology <EG2>
F;197-455/Product: protein C heavy chain #status experimental <HCH>
F;197-210/Domain: activation peptide #status experimental <APT>
F;211-440/Domain: trypsin homology <TRY>
F;45,46,53,55,58,59,62,64,65,68,74/Modified site: gamma-carboxyglutamic acid (Glu) #stat
F;110/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental
F;119-128,137-148,144-157,159-172,180-318,237-253,368-382,393-421/Disulfide bonds: #stat
F;136,289,350/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;252,298,397/Active site: His, Asp, Ser #status predicted
F;366/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.9%; Score 80.5; DB 1; Length 456;
Best Local Similarity 24.8%; Pred. No. 2.5;

Matches 26; Conservative 10; Mismatches 40; Indels 29; Gaps 4;

```
Qy   18 SKTLCSMEEAINERIQEVAGSLIFRAISSIGLECQSVTSRG--DLATCPRGFAVTG---- 72
            |:|:|| ||| |:||:|| :|:||| :|||:|| ||:|||
Db   57 SEEVCEFEEA-REIFQNTEDTMAFWSFYSDGDQCEDRPSGSPCDLPCCGRGKCIDGLGGF 115

Qy   72 -CTC-----------------GSACGSWDVRAETTCHCQCA 94
                             |:||
Db  116 RCDCAEGWEGRFCLHEVRFSNCSAENGGCAHYCMEEEGRRHCSCA 160
```

RESULT 6
T18739

hypothetical protein B0393.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18739
R;Sulston, J.
submitted to the EMBL Data Library, September 1994
A;Reference number: Z19013
A;Accession: T18739
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1106 <WII>
A;Cross-references: EMBL:Z37983; NID:e1519039; PIDN:CAA86058.1; GSPDB:GN00021; CESP:B
A;Experimental source: clone B0393
C;Genetics:
A;Gene: CESP:B0393.5
A;Map position: 3
A;Introns: 15/3; 60/1; 100/3; 343/2; 443/1; 507/3; 609/3; 659/1; 796/1; 876/1; 945/1;

Query Match 13.4%; Score 78; DB 2; Length 1106;
Best Local Similarity 31.1%; Pred. No. 9;
Matches 19; Conservative 4; Mismatches 22; Indels 16; Gaps 3; -

```
Qy   59 DLATCPRG-FAVTGCTCGSACGSWDVRAETTCHC-------------QCAG---MDWTGAR 102
            || |||||| :|:||| |||||||:|:|||||           :||:|
Db  772 DLQTCPPGRLGEPLCDRECAAGHYGINCESTCHCDGSVACDVITGMCPGALCRAGWEGSS 831

Qy  103 C 103
           |
Db  832 C 832
```

RESULT 7
JC5559

lectin-B - Virginian pokeweed
C;Species: Phytolacca americana (Virginian pokeweed)
C;Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 12-Dec-1997
C;Accession: JC5559
R;Yamaguchi, K.; Yurino, N.; Kino, M.; Ishiguro, M.; Funatsu, G.
Biosci. Biotechnol. Biochem. 61, 690-698, 1997
A;Title: The amino acid sequence of mitogenic lectin-B from the roots of pokeweed (Ph
A;Reference number: JC5559; MUID:97290889
A;Accession: JC5559
A;Molecule type: protein
A;Residues: 1-295 <YAM>
A;Experimental source: root
C;Comment: This protein is a lectin specific for N-acetylgucosamine-containing saccha
C;Keywords: glycoprotein
F;96,139/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.3%; Score 77; DB 2; Length 295;
Best Local Similarity 34.6%; Pred. No. 3.7;
Matches 18; Conservative 4; Mismatches 18; Indels 12; Gaps 3;

```
Qy   62 TCPRGFAVTGCTCGSACGSWDVRAETTCHCQC----AGMDWTG-----ARCC 104
           ||| ||||| || |:|| | ||||||||||    :|:||     |||
Db  223 TCPNELC---CSSGGWCGSNDAHCGKGCGQSQCDYWRCGVDFSGRVCPQGRCC 271
```

RESULT 8
B49175
Motch A protein - mouse (fragment)
N;Alternate names: Notch homolog
C;Species: Mus musculus (house mouse)
C;Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 20-Sep-1999
C;Accession: B49175; PHI569; S32109
R;Lardelli, M.; Lendahl, U.
Exp.Cell Res. 204, 364-372, 1993
A;Title: Motch A and Motch B--two mouse Notch homologues coexpressed in a wide variety o
A;Reference number: A49175; MUID:93178563
A;Accession: B49175
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-387 <LAR>
A;Cross-references: EMBL:X68278; NID:g287987; PIDN:CAA48339.1; PID:g287988
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIP:126159)
C;Comment: This protein has many EGF repeats and lin-12/Notch repeats.
C;Comment: This protein is one of the neurogenic proteins controlling the decision betwe
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
F;27-58/Domain: EGF homology <EGF>
F;73-104/Domain: EGF homology <EGF2>
F;151-185/Domain: EGF homology <EGF1>

Query Match   13.0%;   Score 75.5; DB 2; Length 387;
Best Local Similarity   29.1%;   Pred. No. 6.5;
Matches 25; Conservative 4; Mismatches 24; Indels 33; Gaps 5;

Qy   48 GLECQSV------------------------TSRGDLATCPRGFAVTGCTC---GSACGSWDVR 84

Db   141 GRRCESVINGCRGKPCKNGGVCAVASNTARGFICRCPAGF--EGATCENDARTCGSLRCL 198

Qy   85 AETTC------HCQCAGMDWTGARC 103

Db   199 NGGTCISGPRSPTCLCLG-SFTGPEC 223

RESULT 9
A35672
crumbs protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 21-Sep-1990 #sequence_revision 18-Nov-1992 #text_change 13-Aug-1999
C;Accession: A35672
R;Tepass, U.; Theres, C.; Knust, E.
Cell 61, 787-799, 1990
A;Title: crumbs encodes an EGF-like protein expressed on apical membranes of Drosophila
A;Reference number: A35672; MUID:90263104
A;Accession: A35672
A;Status: preliminary
A;Molecule type: mRNA
A;Cross-references: GB:M33753
A;Note: the authors translated the codon GGC for residue 1928 as Cys, and TAT for residu
C;Genetics:
A;Gene: FlyBase:crb
A;Cross-references: FlyBase:FBgn0000368
C;Superfamily: EGF homology
C;Keywords: transmembrane protein
F;352-385/Domain: EGF homology <EGX1>
F;392-424/Domain: EGF homology <EGF1>
F;691-722/Domain: EGF homology <EGF>
F;767-799/Domain: EGF homology <EGF3>
F;1878-1914/Domain: EGF homology <EGX2>

Query Match   13.0%;   Score 75.5; DB 2; Length 2139;
Best Local Similarity   25.0%;   Pred. No. 26;
Matches 24; Conservative 10; Mismatches 37; Indels 25; Gaps 4;

Qy   16 VSSKTLCSMEEAINERIQEVAGSLIFRAISSIGLECQSVTSRGDLATCPRGFAVTGCTCG 75

Db   346 VNIHPLCQTNPCLNNGACVVIG------GSGALTCE--------CPKGYAGARCEVD 388

Qy   76 S-ACGSWDVRAETTC------HCQCAGMDWTGARC 103

Db   389 TDECASQPCQNNGSCIDRINGFSCDCSGTGYTGAFC 424

RESULT 10
A46019
Notch-1 protein - mouse
N;Alternate names: motch protein
C;Species: Mus musculus (house mouse)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
C;Accession: A46019; S25144
R;del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; G
Genomics 15, 259-264, 1993
A;Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog
A;Reference number: A46019; MUID:93194170
A;Accession: A46019
A;Status: not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-2531 <DEL>
A;Cross-references: GB:Z11886; GB:S47228; NID:g288502; PIDN:CAA77941.1; PID:g288503
A;Reference number: S25144
R;Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J
submitted to the EMBL Data Library, April 1992
A;Description: Expression pattern of Motch, a mouse homolog of Drosophila Notch, sugg
A;Reference number: S25144
A;Accession: S25144
A;Molecule type: mRNA
A;Residues: 1551-2108,'Q',2110-2114,'ALP',2118-2170 <FRA>
A;Cross-references: EMBL:Z11886
C;Genetics:
A;Gene: notch-1
A;Map position: 2
A;Note: proximal region of chromosome 2
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol
F;106-138/Domain: EGF homology <EGF1>
F;144-175/Domain: EGF homology <EG01>
F;222-254/Domain: EGF homology <EGF2>
F;261-292/Domain: EGF homology <EG02>
F;339-370/Domain: EGF homology <EGF3>
F;416-449/Domain: EGF homology <EG03>
F;456-487/Domain: EGF homology <EGF4>
F;494-525/Domain: EGF homology <EG04>
F;532-563/Domain: EGF homology <EGF5>
F;607-638/Domain: EGF homology <EG05>
F;682-713/Domain: EGF homology <EG06>
F;757-788/Domain: EGF homology <EG07>
F;795-826/Domain: EGF homology <EG08>
F;873-904/Domain: EGF homology <EG09>
F;911-942/Domain: EGF homology <EG10>
F;949-980/Domain: EGF homology <EG11>
F;987-1018/Domain: EGF homology <EG12>
F;1025-1056/Domain: EGF homology <EG13>
F;1063-1094/Domain: EGF homology <EG14>
F;1149-1180/Domain: EGF homology <EG15>
F;1187-1218/Domain: EGF homology <EG16>
F;1233-1264/Domain: EGF homology <EG17>
F;1352-1383/Domain: EGF homology <EG18>
F;1391-1425/Domain: EGF homology <EG19>
F;1917-1948/Domain: ankyrin repeat homology <AN1>
F;1949-1981/Domain: ankyrin repeat homology <AN2>
F;1983-2015/Domain: ankyrin repeat homology <AN3>
F;2016-2048/Domain: ankyrin repeat homology <AN4>
F;2049-2081/Domain: ankyrin repeat homology <AN5>

Query Match   13.0%;   Score 75.5; DB 2; Length 2531;
Best Local Similarity   29.1%;   Pred. No. 30;
Matches 25; Conservative 4; Mismatches 24; Indels 33; Gaps 5;

Qy   48 GLECQSV------------------------TSRGDLATCPRGFAVTGCTC---GSACGSWDVR 84

C;Species: Homo sapiens (man)
C;Date: 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change 13-Aug-1999
C;Accession: A40043
R;Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds, T.C.; Smith, S.D.; Skl
cell 66, 649-661, 1991
A;Title: TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromoso
A;Reference number: A40043; MUID:91347367
A;Accession: A40043
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A;Molecule type: mRNA
A;Residues: 1-2555 <ELL>
A;Cross-references: GB:M73980
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology
F;261-292/Domain: EGF homology <EGX1>
F;494-525/Domain: EGF homology <EGF1>
F;987-1018/Domain: EGF homology <EGF>
F;1149-1180/Domain: EGF homology <EGF>
F;1187-1218/Domain: EGF homology <EGF3>
F;1233-1264/Domain: EGF homology <EGX3>
F;1927-1959/Domain: ankyrin repeat homology <AN1>
F;1960-1992/Domain: ankyrin repeat homology <AN2>
F;1994-2026/Domain: ankyrin repeat homology <AN3>
F;2027-2059/Domain: ankyrin repeat homology <AN4>
F;2060-2092/Domain: ankyrin repeat homology <AN5>

Query Match   12.8%;   Score 74.5;   DB 2;   Length 2555;
Best Local Similarity   29.1%;   Pred. No. 38;
Matches   25;   Conservative   4;   Mismatches   24;   Indels   33;   Gaps   5;

Qy   48 GLECQSV------------------------------TSRGDLATCPRGFAVTGCTC---GSACGSWDVR 84
Db   1301 GRRCESVINGCKGKPCKNGGTCAVASNTARGFICKCPAGF--EGATCENDARTCGSLRCL 1358

Qy   85 AETTC-------HCCQAGMDWTGARC 103
Db   1359 NGGTCISGPRSPTCLCLG-PFTGPEC 1383

RESULT 14
S18188
notch protein homolog - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 13-Aug-1999
C;Accession: S18188
R;Weinmaster, G.; Roberts, V.J.; Lemke, G.
Development 113, 199-205, 1991
A;Title: A homolog of Drosophila Notch expressed during mammalian development.
A;Reference number: S18188; MUID:92111383
A;Accession: S18188
A;Molecule type: mRNA
A;Residues: 1-2531 <WEI>
A;Cross-references: EMBL:X57405; NID:g57634; PID:g57635
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology
F;987-1018/Domain: EGF homology <EGF1>
F;1025-1056/Domain: EGF homology <EGF>
F;1233-1264/Domain: EGF homology <EGF2>
F;1917-1949/Domain: ankyrin repeat homology <AN1>
F;1950-1982/Domain: ankyrin repeat homology <AN2>
F;1988-2016/Domain: ankyrin repeat homology <AN3>
F;2017-2049/Domain: ankyrin repeat homology <AN4>
F;2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match   12.7%;   Score 74;   DB 2;   Length 2531;
Best Local Similarity   25.7%;   Pred. No. 42;
Matches   36;   Conservative   12;   Mismatches   52;   Indels   40;   Gaps   8;

Qy   4 LCLLLPVL--GLLVSSKTLCSMEEAINERIQEVAGSLIFRAISS--IGLECQS------ 54
Db   9 LCLTLLPALAARGLRCSQPS---GTCLNGGRCEVANGTEACVCSGAFVGQRCQDPSPCL 64
Qy   54 ----------VTSRGDL----ATCPRGFAVTGC-TCGSAC-------GSWDVRAETT 88

Db   1301 GRRCESVINGCRGKPCKNGGVCAVASNTARGFICRCPAGF--EGATCENDARTCGSLRCL 1358

Qy   85 AETTC-------HCCQAGMDWTGARC 103
Db   1359 NGGTCISGPRSPTCLCLG-SFTGPEC 1383

RESULT 11
A21761
high-cysteine chorion A 12 protein precursor - silkworm
C;Species: Bombyx mori (silkworm)
C;Date: 20-Jul-1990 #sequence_revision 20-Jul-1990 #text_change 26-Aug-1999
C;Accession: A21761
R;Iatrou, K.; Tsitilou, S.G.; Kafatos, F.C.
Proc. Natl. Acad. Sci. U.S.A. 81, 4452-4456, 1984
A;Title: DNA sequence transfer between two high-cysteine chorion gene families in the si
A;Reference number: A21761; MUID:84272653
A;Accession: A21761
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-124 <IAT>
C;Superfamily: chorion class A protein pc292

Query Match   12.9%;   Score 75;   DB 2;   Length 124;
Best Local Similarity   28.7%;   Pred. No. 2.8;
Matches   29;   Conservative   7;   Mismatches   31;   Indels   34;   Gaps   8;

Qy   9 LPVLGLLVSSKTLCSMEEAINERIQEVAGSLIFRAISSIGLECQSVTSRGDLATCPR--G 66
Db   51 LPVCG------------ETVVCGRV-PICGGVCFK----GPACAS---GCVSICGRCCG 89
Qy   67 FAVTGC-TCGSACGSWDVRAETTCHCCQAGMDWTGA--RCC 104
Db   90 CGCGGCGGCGCGCGG--------CGCGCGGCGGCGGRRCC 122

RESULT 12
T37097
probable secreted neuraminidase - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T37097
R;Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, August 1999
A;Reference number: 221588
A;Accession: T37097
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-748 <SAU>
A;Cross-references: EMBL:AL109950; PIDN:CAB52948.1; GSPDB:GN00070; SCOEDB:SCJ4.14c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCJ4.14c

Query Match   12.9%;   Score 75;   DB 2;   Length 748;
Best Local Similarity   29.3%;   Pred. No. 12;
Matches   24;   Conservative   12;   Mismatches   28;   Indels   18;   Gaps   3;

Qy   13 GLLVSSKTLCSMEEAINERIQE-------VAGSLIFRAISSIGLECQSVTSRGDLATCPRG 66
Db   468 GTGIAGKTV-----AFNDAVQQHVLARGTDGSLIHRFWSTTGLSSQTWLAAGTVAGAPTG 522
Qy   67 FAVTG------CTCGSACGSW 81
Db   523 FVASGQQHAWVRMTDGRLLHTW 544

RESULT 13
A40043
notch protein homolog TAN-1 precursor - human

```
Db    65 STPCKNAGTCYVVDHGGIVDYACSCPLGFSGPLCLTPLANACLANPCRNGGTCDLLTLTE 124
          |   :   :|| ||:   |   :||  |   |:   |: ||:   |
Qy    89 CHCQCAGMDWTGARCCRVQP 108
          |:|   |:|   |   :   |
Db   125 YKCRCP-PGWSGKSCQQADP 143


RESULT  15
T25682
hypothetical protein F08D12.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T25682
R;Le, T.; Waterston, R.
submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid F08D12.
A;Reference number: Z20068
A;Accession: T25682
A;Status: preliminary: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-289 <LET>
A;Cross-references: EMBL:U80840; PIDN:AAB37924.1; GSPDB:GN00020; CESP:F08D12.3
A;Experimental source: strain Bristol N2; clone F08D12
C;Genetics:
A;Gene: CESP:F08D12.3
A;Map position: 2
A;Introns: 17/1; 56/1; 93/1; 134/1; 171/1; 213/1; 248/1

Query Match         12.7%;  Score 73.5;  DB 2;  Length 289;
Best Local Similarity  28.7%;  Pred. No. 7.8;
Matches   27;  Conservative  9;  Mismatches  37;  Indels  21;  Gaps  4;

Qy    13 GLLVSSKTLCSMEEAINERIQEVAGSLIFRAISSIGLECQSVTSRGDLATCPRGFAVTGC 72
          |   :   ::|   ||   ||:|   |               |:|   |   ||   |   |:   |
Db   117 GASITKGSVCCW---NEGIQDAPAS----------ACKSSASPAILGICPSGQVLIGN 161

Qy    73 TCGSACGSWDVRAETTCHCQC---AGMDWTGARC 103
          |   |:||   :||   :||   ||   ||   |   |
Db   162 YC---CEAKDVYDPSTATCQRGYEAGPYGTGLTC 192


Search completed: May 26, 2000, 03:18:28
Job time: 10177 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2000, 03:01:52 ; Search time 40.45 Seconds
(without alignments)
81.314 Million cell updates/sec

Title: US-09-099-898-2
Perfect score: 581
Sequence: 1 MKALCLLLLPVLGLLVSSKT.........CHCQCAGMDWTGARCCRVQP 108

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :   SwissProt_38:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	86	14.8	934	1	CO6_HUMAN	P13671 homo sapien
2	80.5	13.9	456	1	PRTC_BOVIN	P00745 bos taurus
3	75.5	13.0	2139	1	CRB_DROME	P10040 drosophila
4	75.5	13.0	2531	1	NTC1_MOUSE	Q01705 mus musculu
5	74.5	12.8	2444	1	NTC1_HUMAN	P46531 homo sapien
6	74	12.7	124	1	CHH2_BOMMO	P05687 bombyx mori
7	74	12.7	2531	1	NTC1_RAT	Q07008 rattus norv
8	73.5	12.7	660	1	PRIA_BORBU	Q45032 borrelia bu
9	73	12.6	1049	1	ADP1_YEAST	P25371 saccharomyc
10	73	12.6	1700	1	BAR3_CHITE	Q03376 chironomus
11	73	12.6	2437	1	NOTC_BRARE	P46530 brachydanio
12	72	12.4	1193	1	LMG2_HUMAN	Q13753 homo sapien
13	72	12.4	2318	1	NTC3_MOUSE	Q61982 mus musculu
14	71.5	12.3	519	1	TYR2_HUMAN	P40126 homo sapien
15	71.5	12.3	2476	1	ZAN_PIG	Q28983 sus scrofa
16	71.5	12.3	2703	1	NOTC_DROME	P07207 drosophila
17	71	12.2	72	1	MT13_MYTED	P80248 mytilus edu
18	71	12.2	397	1	PRIA_LYSEN	P00778 lysobacter
19	71	12.2	1134	1	TIE1_MOUSE	Q06806 mus musculu
20	71	12.2	2907	1	FBN2_MOUSE	Q61555 mus musculu
21	70	12.0	2871	1	FBN1_MOUSE	Q61554 mus musculu
22	70	12.0	2524	1	NOTC_XENLA	Q21783 xenopus lae
23	69.5	11.9	2715	1	G156_PARPR	P13837 paramecium
24	68.5	11.8	581	1	IRR_RAT	Q64716 rattus norv
25	68	11.7	72	1	MT11_MYTED	P80246 mytilus edu
26	68	11.7	458	1	PRTC_RABIT	Q28661 oryctolagus
27	68	11.7	1138	1	TIE1_HUMAN	P35590 homo sapien
28	68	11.7	72	1	MT12_MYTED	P80247 mytilus edu
29	67	11.5	364	1	PGLR_COCCA	P26215 cochliobolu
30	67	11.5	372	1	AG1_URTDI	P11218 urtica diol
31	67	11.5	1136	1	TIE1_BOVIN	Q06805 bos taurus
32	67	11.5	1801	1	LMB2_RAT	P15800 rattus norv
33	66.5	11.4	71	1	MT21_MYTED	P80251 mytilus edu
34	66.5	11.4	182	1	HPRT_RHOCA	P37171 rhodobacter
35	66.5	11.4	444	1	SHU4_ECOLI	P09748 escherichia
36	66.5	11.4	1416	1	YN81_CAEEL	Q03610 caenorhabdi
37	66.5	11.4	2911	1	FBN2_HUMAN	P35556 homo sapien
38	66	11.4	72	1	MT14_MYTED	P80249 mytilus edu
39	66	11.4	177	1	BTC_MOUSE	Q05928 mus musculu
40	66	11.4	309	1	PNAD_PIG	Q28955 sus scrofa
41	66	11.4	463	1	MFGM_MOUSE	P21956 mus musculu
42	66	11.4	1300	1	IRR_CAVPO	P14617 cavia porce
43	66	11.4	2871	1	FBN1_HUMAN	P35555 homo sapien
44	65.5	11.3	66	1	MT_ARIAR	P55946 arianta arb
45	65.5	11.3	71	1	MT22_MYTED	P80252 mytilus edu

ALIGNMENTS

RESULT 1
CO6_HUMAN
ID   CO6_HUMAN   STANDARD;   PRT;   934 AA.
AC   P13371;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   COMPLEMENT COMPONENT C6 PRECURSOR.
GN   C6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 22-31 AND 633-640.
RX   MEDLINE; 90036879.
RA   Discipio R.G., Hugli T.E.;
RT   "The molecular architecture of human complement component C6.";
RL   J. Biol. Chem. 264:16197-16205(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89380223.
RA   Haefliger J.-A., Tschopp J., Vial N., Jenne D.E.;
RT   "Complete primary structure and functional characterization of the
RT   sixth component of the human complement system. Identification of the
RT   C5b-binding domain in complement C6.";
RL   J. Biol. Chem. 264:18041-18051(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE; 93291175.
RA   Hobart M.J., Fernie B., Disciplo R.G.;
RT   "Structure of the human C6 gene.";
RL   Biochemistry 32:6198-6205(1993).
RN   [4]
RP   SEQUENCE OF 1-491 FROM N.A.
RX   MEDLINE; 89202413.
RA   Chakravarti D.N., Chakravarti B., Parra C.A., Mueller-Eberhard H.J.;
RT   "Structural homology of complement protein C6 with other
RT   channel-forming proteins of complement.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:2799-2803(1989).
RN   [5]
RP   VARIANT ALLOTYPES C6 A AND B.
RX   MEDLINE; 93326158.
RA   Dewald G., Nothen M.M., Cichon S.;
RT   "Polymorphism of human complement component C6: an amino acid
RT   substitution (Glu/Ala) within the second thrombospondin repeat
RT   differentiates between the two common allotypes C6 A and C6 B.";
RL   Biochem. Biophys. Res. Commun. 194:458-464(1993).
RN   [6]
RP   CARBOHYDRATE-BINDING SITES.
RX   MEDLINE; 20020247.
RA   Hofsteenge J., Blommers M., Hess D., Furmanek A., Miroshnichenko O.;
RT   "The four terminal components of the complement system are
RT   C-mannosylated on multiple tryptophan residues.";
RL   J. Biol. Chem. 274:32786-32794(1999).
CC   -!- FUNCTION: INVOLVED IN THE FORMATION OF THE LYTIC C5b-9M COMPLEX.
CC   -!- DOMAIN: C6 CONTAINS NINE DISTINCT STRUCTURAL UNITS (CYSTEINE-RICH)

```
CC        WHICH HAVE THE CHARACTERISTIC FEATURES OF MODULES.
CC   -!- PTM: ALL CYSTEINE RESIDUES ARE ASSUMED TO BE CROSS-LINKED TO ONE
CC        ANOTHER. INDIVIDUAL MODULES CONTAINING AN EVEN NUMBER OF CONSERVED
CC        CYSTEINE RESIDUES ARE SUPPOSED TO HAVE DISULFIDE LINKAGES ONLY
CC        WITHIN THE SAME MODULE.
CC   -!- POLYMORPHISM: THE SEQUENCE SHOWN IS THAT OF ALLOTYPE C6 A.
CC   -!- SIMILARITY: TO COMPLEMENT FACTORS C7, C8, C9, AND TO PERFORIN.
CC   -!- SIMILARITY: CONTAINS 2 SUSHI (SCR) REPEATS.
CC   -!- SIMILARITY: CONTAINS 3 TYPE-1 TSP REPEATS.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC   --------------------------------------------------------------------------
CC        This SWISS-PROT entry is copyright. It is produced through a collaboration
CC        between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC        the European Bioinformatics Institute. There are no restrictions on its
CC        use by non profit institutions as long as its content is in no way
CC        modified and this statement is not removed. Usage by and for commercial
CC        entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC        or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; J05064; AAA51860.1; -.
DR   EMBL; J05024; AAA59668.1; -.
DR   EMBL; X72177; CAA50994.1; -.
DR   EMBL; J04506; AAB59433.1; -.
DR   PIR; A32109; A32109.
DR   PIR; A34235; A34235.
DR   PIR; A34372; A34372.
DR   HSSP; P01130; 1AJJ.
DR   MIM; 217050; -.
DR   PRINTS; PR00764; COMPLEMENTC9.
DR   PROSITE; PR00279; MAC_PERFORIN; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS01209; LDLRA_1; 1.
DR   PROSITE; PS50068; LDLRA_2; 1.
DR   PFAM; PF00050; kazal; 1.
DR   PFAM; PF00057; ldl_recept_a; 1.
DR   PFAM; PF00084; sushi; 2.
DR   PFAM; PF00090; tsp_1; 3.
KW   Complement pathway; Glycoprotein; Plasma; Membrane attack complex;
KW   Cytolysis; Sushi; Repeat; Signal; Polymorphism; EGF-like domain;
KW   Transmembrane.
FT   SIGNAL        1     21
FT   CHAIN        22    934       COMPLEMENT COMPONENT C6.
FT   REPEAT       22     80       TYPE-1 TSP 1.
FT   REPEAT       81    137       TYPE-1 TSP 2.
FT   DOMAIN      138    175       LDL-RECEPTOR CLASS A.
FT   TRANSMEM    331    349       POTENTIAL.
FT   TRANSMEM    354    373       POTENTIAL.
FT   DOMAIN      517    553       EGF-LIKE.
FT   REPEAT      562    611       TYPE-1 TSP 3.
FT   DOMAIN      642    934       C5B-BINDING DOMAIN.
FT   DOMAIN      643    762       2 X SUSHI (SCR) REPEATS.
FT   REPEAT      643    700       SUSHI 1.
FT   REPEAT      703    762       SUSHI 2.
FT   DOMAIN      766    840       COMPLEMENT CONTROL FACTOR I MODULE.
FT   DOMAIN      858    934       COMPLEMENT CONTROL FACTOR I MODULE.
FT   DISULFID    140    151       BY SIMILARITY.
FT   DISULFID    146    164       BY SIMILARITY.
FT   DISULFID    158    173       BY SIMILARITY.
FT   DISULFID    399    420       BY SIMILARITY.
FT   DISULFID    523    539       BY SIMILARITY.
FT   DISULFID    526    541       BY SIMILARITY.
FT   DISULFID    543    552       BY SIMILARITY.
FT   DISULFID    644    686       BY SIMILARITY.
FT   DISULFID    672    699       BY SIMILARITY.
FT   DISULFID    704    746       BY SIMILARITY.
FT   DISULFID    732    761       BY SIMILARITY.
FT   CARBOHYD     29     29       C-LINKED (MAN) (PARTIAL).
FT   CARBOHYD     32     32       C-LINKED (MAN) (PARTIAL).
FT   CARBOHYD     90     90       C-LINKED (MAN) (PARTIAL).
FT   CARBOHYD    324    324       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    568    568       C-LINKED (MAN) (PARTIAL).
```

```
FT   CARBOHYD    571    571       C-LINKED (MAN) (PARTIAL).
FT   CARBOHYD    574    574       C-LINKED (MAN) (PARTIAL).
FT   CARBOHYD    855    855.      N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARIANT     119    119       E -> A (IN ALLOTYPE C6 B).
                                  /FTId=VAR_006056.
SQ   SEQUENCE   934 AA;  104843 MW;  5A2713AB5B507B80 CRC64;

Query Match             14.8%;  Score 86;  DB 1;  Length 934;
Best Local Similarity   29.4%;  Pred. No. 0.33;
Matches   30;  Conservative 10;  Mismatches 36;  Indels 26;  Gaps 7;

Qy     22 CSMEEAINERIQEVAGSLIFRAISSIGLECQSVTSRGDLATCPRGFAVTG---CTCGSAC 78
          ||  :|                  ||  :|           ||:|:||  ||    |:
Db    699 CQRTECIKPVVQEVLTITPFQRLYRIG---ESIE-----LTCPKGFVAGPSRYTCQG-- 749

Qy     79 GSW--DVRAETTC----------HCQCAGMDWTGARCCRVQP 108
          ||  |                  |||  |  | |:  :  :
Db    749 NSWTPPISNSLTCEKDTLTKLKGHCQ-LGQKQSGSECICMSP 789


RESULT   2
PRTC_BOVIN
ID   PRTC_BOVIN     STANDARD;      PRT;     456 AA.
AC   P00745;
DT   21-JUL-1986 (Rel. 01, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   VITAMIN-K DEPENDENT PROTEIN C PRECURSOR (EC 3.4.21.69)
DE   (AUTOPROTHROMBIN IIA) (ANTICOAGULANT PROTEIN C) (BLOOD COAGULATION
DE   FACTOR XIV) (FRAGMENT).
GN   PROC.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85014826.
RA   Long G.L., Balagaje R.M., McGillivray R.T.A.;
RT   "Cloning and sequencing of liver cDNA coding for bovine protein C.";
RL   Proc. Natl. Acad. Sci. U.S.A. 81:5653-5656(1984).
RN   [2]
RP   SEQUENCE OF 40-194.
RX   MEDLINE; 83007325.
RA   Fernlund P., Stenflo J.;
RT   "Amino acid sequence of the light chain of bovine protein C.";
RL   J. Biol. Chem. 257:12170-12179(1982).
RN   [3]
RP   REVISION TO 110.
RX   MEDLINE; 83169769.
RA   Drakenberg T., Fernlund P., Roepstorff P., Stenflo J.;
RT   "Beta-hydroxyaspartic acid in vitamin K-dependent protein C.";
RL   Proc. Natl. Acad. Sci. U.S.A. 80:1802-1806(1983).
RN   [4]
RP   SEQUENCE OF 197-456.
RX   MEDLINE; 83007326.
RA   Stenflo J., Fernlund P.;
RT   "Amino acid sequence of the heavy chain of bovine protein C.";
RL   J. Biol. Chem. 257:12180-12190(1982).
RN   [5]
RP   PROCESSING, AND CALCIUM-BINDING DATA.
RX   MEDLINE; 83213513.
RA   Esmon N.L., Debault L.E., Esmon C.T.;
RT   "Proteolytic formation and properties of gamma-carboxyglutamic acid-
RT   domainless protein C.";
RL   J. Biol. Chem. 258:5548-5553(1983).
RN   [6]
RP   PROCESSING, AND CALCIUM-BINDING DATA.
RX   MEDLINE; 83213514.
RA   Johnson A.E., Esmon N.L., Laue T.M., Esmon C.T.;
RT   "Structural changes required for activation of protein C are induced
RT   by Ca2+ binding to a high affinity site that does not contain gamma-
```

```
RT   carboxyglutamic acid.";
RL   J. Biol. Chem. 258:5554-5560(1983).
CC   -!- FUNCTION: PROTEIN C IS A VITAMIN K-DEPENDENT SERINE PROTEASE THAT
CC       REGULATES BLOOD COAGULATION BY INACTIVATING FACTORS VA AND VIIIA
CC       IN THE PRESENCE OF CALCIUM IONS AND PHOSPHOLIPIDS.
CC   -!- CATALYTIC ACTIVITY: DEGRADATION OF BLOOD COAGULATION FACTORS VA
CC       AND VIIIA.
CC   -!- SUBUNIT: SYNTHESIZED AS A SINGLE CHAIN PRECURSOR, WHICH IS CLEAVED
CC       INTO A LIGHT CHAIN AND A HEAVY CHAIN HELD TOGETHER BY A DISULFIDE
CC       BOND. THE ENZYME IS THEN ACTIVATED BY THROMBIN, WHICH CLEAVES A
CC       TETRADECAPEPTIDE FROM THE AMINO END OF THE HEAVY CHAIN; THIS
CC       REACTION, WHICH OCCURS AT THE SURFACE OF ENDOTHELIAL CELLS, IS
CC       STRONGLY PROMOTED BY THROMBOMODULIN.
CC   -!- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED IN THE LIVER.
CC   -!- PTM: THE VITAMIN K-DEPENDENT, ENZYMATIC CARBOXYLATION OF SOME
CC       GLU RESIDUES ALLOWS THE MODIFIED PROTEIN TO BIND CALCIUM.
CC   -!- MISCELLANEOUS: CALCIUM ALSO BINDS, WITH STRONGER AFFINITY TO
CC       ANOTHER SITE, BEYOND THE GLA DOMAIN. THIS GLA-INDEPENDENT BINDING
CC       SITE IS NECESSARY FOR THE RECOGNITION OF THE
CC       THROMBIN-THROMBOMODULIN COMPLEX.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   --------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; K02435; AAA30685.1; -.
DR   PIR; A00928; KXBO.
HSSP; P04070; 1AUT.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00011; GLU_CARBOXYLATION; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; FALSE_NEG.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PFAM; PF00008; EGF; 2.
DR   PFAM; PF00089; trypsin; 1.
DR   PFAM; PF00594; gla; 1.
KW   Blood coagulation; Glycoprotein; Plasma; Serine protease; Liver;
KW   Gamma-carboxyglutamic acid; Calcium-binding; Vitamin K; Hydroxylation;
KW   EGF-like domain; Repeat; Endothelial cell; Hydrolase; Signal.
FT   NON_TER        1      1
FT   SIGNAL        <1     29
FT   PROPEP        30     39
FT   CHAIN         40    194       PROTEIN C LIGHT CHAIN.
FT   CHAIN        197    456       PROTEIN C HEAVY CHAIN.
FT   PEPTIDE      197    210       ACTIVATION PEPTIDE.
FT   DOMAIN        94    129       EGF-LIKE 1.
FT   DOMAIN       133    173       EGF-LIKE 2.
FT   DOMAIN       211    456       CATALYTIC.
FT   MOD_RES       45     45       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES       46     46       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES       53     53       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES       55     55       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES       58     58       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES       59     59       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES       62     62       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES       64     64       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES       65     65       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES       68     68       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES       74     74       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      110    110       HYDROXYLATION.
FT   ACT_SITE     252    252       CHARGE RELAY SYSTEM.
FT   ACT_SITE     298    298       CHARGE RELAY SYSTEM.
FT   ACT_SITE     397    397       CHARGE RELAY SYSTEM.
FT   DISULFID      56     61       BY SIMILARITY.
FT   DISULFID      89    108       BY SIMILARITY.
FT   DISULFID      98    103       BY SIMILARITY.
FT   DISULFID     102    117       BY SIMILARITY.
FT   DISULFID     119    128       BY SIMILARITY.
FT   DISULFID     137    148       BY SIMILARITY.
FT   DISULFID     144    157       BY SIMILARITY.
FT   DISULFID     159    172       BY SIMILARITY.
FT   DISULFID     180    318       INTERCHAIN.
FT   DISULFID     237    253
FT   DISULFID     368    382
FT   DISULFID     393    421
FT   CARBOHYD     136    136
FT   CARBOHYD     289    289
FT   CARBOHYD     350    350
FT   CARBOHYD     366    366       ASN-X-CYS (INSTEAD OF SER/THR).
FT   VARIANT       82     82       F -> K.
FT   CONFLICT     455    456       VP -> PV (IN REF. 4).
SQ   SEQUENCE     456 AA;   51407 MW;  CAAF6833F894C209 CRC64;

Query Match             13.9%;  Score 80.5;  DB 1;  Length 456;
Best Local Similarity   24.8%;  Pred. No.0.61;
  Matches   26;  Conservative  10;  Mismatches  40;  Indels  29;  Gaps   4;

Qy       18 SKTLCSMEEAINERIQEVAGSLIFRAISSIGLECQSVTSRG--DLATCPRGFAVTG---- 72
            | :| :|| |  :|  :| |  || : | |:|::|:   :|:    :   |  |    |
Db       57 SEEVCEFEEA-REIFQNTEDTMAFWSFYSDGQCEDRPSGSPCDLPCCGRGKCIDGLGGF 115

Qy       72 -CTC---------------GSACGSWDVRAETTCHCQCA 94
               |                  | :: |  || :||
Db      116 RCDCAEGWEGRFCLHEVRFSNCSAENGGCAHYCMEEGRRHCSCA 160


RESULT  3
CRB_DROME
ID   CRB_DROME       STANDARD;        PRT;   2139 AA.
AC   P10040;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CRUMBS PROTEIN PRECURSOR (95F).
GN   CRB.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=EMBRYO;
RX   MEDLINE; 90263104.
RA   Tepass U., Theres C., Knust E.;
RT   "Crumbs encodes an EGF-like protein expressed on apical membranes of
RT   Drosophila epithelial cells and required for organization of
RT   epithelia.";
RL   Cell 61:787-799(1990).
RN   [2]
RP   SEQUENCE OF 1663-1955 FROM N.A.
RX   MEDLINE; 87218537.
RA   Knust E., Dietrich U., Tepass U., Bremer K.A., Welgel D.,
RA   Vaessin H., Campos-Ortega J.A.;
RT   "EGF homologous sequences encoded in the genome of Drosophila
RT   melanogaster; and their relation to neurogenic genes.";
RL   EMBO J. 6:761-766(1987).
CC   -!- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
CC       POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
CC       POLARITY. IT MAY ACT AS A SIGNAL.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC   --------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
```

```
DR   EMBL; M33753; AAA28428.1; ALT_SEQ.
DR   EMBL; X05144; CAA28793.1; -.
DR   PIR; B26637; B26637.
DR   PIR; A35672; A35672.
DR   HSSP; P00740; 1IXA.
DR   FLYBASE; FBgn0000368; crb.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PROSITE; PS00010; ASX_HYDROXYL; 15.
DR   PROSITE; PS00022; EGF_1; 26.
DR   PROSITE; PS01186; EGF_2; 17.
DR   PROSITE; PS01187; EGF_CA; 15.
DR   PFAM; PF00008; EGF; 27.
DR   PFAM; PF00054; laminin_G; 3.
KW   Differentiation; Repeat; EGF-like domain; Transmembrane;
KW   Glycoprotein; Signal; Phosphorylation.
FT   SIGNAL          1     90
FT   CHAIN          91   2139       CRUMBS PROTEIN.
FT   TRANSMEM     2085   2111       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN       2112   2139       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        267    303       EGF-LIKE 1.
FT   DOMAIN        306    343       EGF-LIKE 2.
FT   DOMAIN        348    386       EGF-LIKE 3.
FT   DOMAIN        388    425       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN        427    463       EGF-LIKE 5.
FT   DOMAIN        464    500       EGF-LIKE 6.
FT   DOMAIN        501    532       EGF-LIKE 7.
FT   DOMAIN        545    581       EGF-LIKE 8.
FT   DOMAIN        582    611       EGF-LIKE 9.
FT   DOMAIN        609    646       EGF-LIKE 10.
FT   DOMAIN        648    685       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN        687    723       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN        725    761       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN        763    800       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN        802    838       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN        840    902       EGF-LIKE 16.
FT   DOMAIN        904    940       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN        942    978       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN        980   1021       EGF-LIKE 19.
FT   DOMAIN       1207   1243       EGF-LIKE 20.
FT   DOMAIN       1481   1517       EGF-LIKE 21.
FT   DOMAIN       1759   1795       EGF-LIKE 22.
FT   DOMAIN       1797   1833       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       1835   1871       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       1874   1915       EGF-LIKE 25.
FT   DOMAIN       1915   1951       EGF-LIKE 26.
FT   DOMAIN       1953   1989       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       1991   2029       EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       2030   2070       EGF-LIKE 29.
FT   DISULFID      271    282       BY SIMILARITY.
FT   DISULFID      276    291       BY SIMILARITY.
FT   DISULFID      293    302       BY SIMILARITY.
FT   DISULFID      310    321       BY SIMILARITY.
FT   DISULFID      315    331       BY SIMILARITY.
FT   DISULFID      333    342       BY SIMILARITY.
FT   DISULFID      352    363       BY SIMILARITY.
FT   DISULFID      357    374       BY SIMILARITY.
FT   DISULFID      376    385       BY SIMILARITY.
FT   DISULFID      392    403       BY SIMILARITY.
FT   DISULFID      397    412       BY SIMILARITY.
FT   DISULFID      414    424       BY SIMILARITY.
FT   DISULFID      431    442       BY SIMILARITY.
FT   DISULFID      436    451       BY SIMILARITY.
FT   DISULFID      453    462       BY SIMILARITY.
FT   DISULFID      468    479       BY SIMILARITY.
FT   DISULFID      473    488       BY SIMILARITY.
FT   DISULFID      490    499       BY SIMILARITY.
FT   DISULFID      505    515       BY SIMILARITY.
FT   DISULFID      509    520       BY SIMILARITY.
FT   DISULFID      522    531       BY SIMILARITY.
FT   DISULFID      549    562       BY SIMILARITY.
FT   DISULFID      556    569       BY SIMILARITY.
FT   DISULFID      571    580       BY SIMILARITY.
FT   DISULFID      586    597       BY SIMILARITY.
FT   DISULFID      591    602       BY SIMILARITY.
FT   DISULFID      604    610       BY SIMILARITY.
FT   DISULFID      613    624       BY SIMILARITY.
FT   DISULFID      618    634       BY SIMILARITY.
FT   DISULFID      636    645       BY SIMILARITY.
FT   DISULFID      652    664       BY SIMILARITY.
FT   DISULFID      659    673       BY SIMILARITY.
FT   DISULFID      675    684       BY SIMILARITY.
FT   DISULFID      691    702       BY SIMILARITY.
FT   DISULFID      696    711       BY SIMILARITY.
FT   DISULFID      713    722       BY SIMILARITY.
FT   DISULFID      729    740       BY SIMILARITY.
FT   DISULFID      734    749       BY SIMILARITY.
FT   DISULFID      751    760       BY SIMILARITY.
FT   DISULFID      767    778       BY SIMILARITY.
FT   DISULFID      772    787       BY SIMILARITY.
FT   DISULFID      789    799       BY SIMILARITY.
FT   DISULFID      806    817       BY SIMILARITY.
FT   DISULFID      811    826       BY SIMILARITY.
FT   DISULFID      828    837       BY SIMILARITY.
FT   DISULFID      844    855       BY SIMILARITY.
FT   DISULFID      849    890       BY SIMILARITY.
FT   DISULFID      892    901       BY SIMILARITY.
FT   DISULFID      908    919       BY SIMILARITY.
FT   DISULFID      913    928       BY SIMILARITY.
FT   DISULFID      930    939       BY SIMILARITY.
FT   DISULFID      946    957       BY SIMILARITY.
FT   DISULFID      952    966       BY SIMILARITY.
FT   DISULFID      968    977       BY SIMILARITY.
FT   DISULFID      984    995       BY SIMILARITY.
FT   DISULFID      989   1009       BY SIMILARITY.
FT   DISULFID     1011   1020       BY SIMILARITY.
FT   DISULFID     1211   1222       BY SIMILARITY.
FT   DISULFID     1216   1231       BY SIMILARITY.
FT   DISULFID     1233   1242       BY SIMILARITY.
FT   DISULFID     1485   1496       BY SIMILARITY.
FT   DISULFID     1490   1505       BY SIMILARITY.
FT   DISULFID     1507   1516       BY SIMILARITY.
FT   DISULFID     1763   1774       BY SIMILARITY.
FT   DISULFID     1768   1783       BY SIMILARITY.
FT   DISULFID     1785   1794       BY SIMILARITY.
FT   DISULFID     1801   1812       BY SIMILARITY.
FT   DISULFID     1806   1821       BY SIMILARITY.
FT   DISULFID     1823   1832       BY SIMILARITY.
FT   DISULFID     1839   1850       BY SIMILARITY.
FT   DISULFID     1844   1859       BY SIMILARITY.
FT   DISULFID     1861   1870       BY SIMILARITY.
FT   DISULFID     1878   1889       BY SIMILARITY.
FT   DISULFID     1883   1903       BY SIMILARITY.
FT   DISULFID     1905   1914       BY SIMILARITY.
FT   DISULFID     1919   1930       BY SIMILARITY.
FT   DISULFID     1924   1939       BY SIMILARITY.
FT   DISULFID     1941   1950       BY SIMILARITY.
FT   DISULFID     1957   1968       BY SIMILARITY.
FT   DISULFID     1962   1977       BY SIMILARITY.
FT   DISULFID     1979   1988       BY SIMILARITY.
FT   DISULFID     1995   2008       BY SIMILARITY.
FT   DISULFID     2002   2017       BY SIMILARITY.
FT   DISULFID     2019   2028       BY SIMILARITY.
FT   CARBOHYD       37     37       POTENTIAL.
FT   CARBOHYD       96     96       POTENTIAL.
FT   CARBOHYD      198    198       POTENTIAL.
FT   CARBOHYD      238    238       POTENTIAL.
FT   CARBOHYD      239    239       POTENTIAL.
FT   CARBOHYD      336    336       POTENTIAL.
FT   CARBOHYD      400    400       POTENTIAL.
FT   CARBOHYD      550    550       POTENTIAL.
```

```
FT   CARBOHYD      565    565       POTENTIAL.
FT   CARBOHYD      736    736       POTENTIAL.
FT   CARBOHYD      746    746       POTENTIAL.
FT   CARBOHYD      860    860       POTENTIAL.
FT   CARBOHYD      884    884       POTENTIAL.
FT   CARBOHYD      976    976       POTENTIAL.
FT   CARBOHYD     1102   1102       POTENTIAL.
FT   CARBOHYD     1114   1114       POTENTIAL.
FT   CARBOHYD     1138   1138       POTENTIAL.
FT   CARBOHYD     1192   1192       POTENTIAL.
FT   CARBOHYD     1245   1245       POTENTIAL.
FT   CARBOHYD     1255   1255       POTENTIAL.
FT   CARBOHYD     1354   1354       POTENTIAL.
FT   CARBOHYD     1363   1363       POTENTIAL.
FT   CARBOHYD     1441   1441       POTENTIAL.
FT   CARBOHYD     1454   1454       POTENTIAL.

Query Match             13.0%;  Score 75.5;  DB 1;  Length 2139;
Best Local Similarity   25.0%;  Pred. No. 7.7;
  Matches   24;  Conservative   10;  Mismatches   37;  Indels   25;  Gaps    4;

Qy     16 VSSKTLCSMEEAINERIQEVAGSLIFRAISSIGLECQSVTSRGDLATCPRGFAVTGCTCG 75
            |:  |:      ||:|                                  ||||
Db    346 VNIHPLCQTNPCLNNGACVVIG------GSGALTCE--------CPKGYAGARCEVD 388

Qy     76 S-ACGSWDVRAETTC------HCQCAGMDWTGARC 103
          :|-|-   :  .| :||   :|||
Db    389 TDECASQPCQNNGSCIDRINGFSCDCSGTGYTGAFC 424

RESULT   4
NTC1_MOUSE
ID   NTC1_MOUSE     STANDARD;      PRT;   2531 AA.
AC   Q01705;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (MOTCH PROTEIN).
GN   NOTCH1 OR MOTCH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93194170.
RA   Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
RA   Copeland N.G., Gridley T.;
RT   "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT   homolog of Drosophila Notch.";
RL   Genomics 15:259-264(1993).
RN   [2]
RP   SEQUENCE OF 1551-2170 FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93048835.
RA   Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
RA   Greenspan R.J., McMahon A.P., Gridley T.;
RT   "Expression pattern of Motch, a mouse homolog of Drosophila Notch,
RT   suggests an important role in early postimplantation mouse
RT   development.";
RL   Development 115:737-744(1992).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
```

```
DR   EMBL; Z11886; CAA77941.1; -.
DR   HSSP; P00740; 1IXA.
DR   MGD; MGI:97363; NOTCH1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
KW   Differentiation; Neurogenesis; Repeat; EGF-like domain;
KW   Transmembrane; Signal; Glycoprotein.
FT   SIGNAL         1     18       POTENTIAL.
FT   CHAIN         19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN        19   1725       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    1726   1746       POTENTIAL.
FT   DOMAIN      1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        20   1426       36 X EGF-TYPE REPEATS.
FT   DOMAIN        20     58       EGF-LIKE 1.
FT   DOMAIN        59     99       EGF-LIKE 2.
FT   DOMAIN       102    139       EGF-LIKE 3.
FT   DOMAIN       140    176       EGF-LIKE 4.
FT   DOMAIN       178    216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       257    293       EGF-LIKE 6.
FT   DOMAIN       295    333       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       335    371       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       372    410       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       412    450       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       452    488       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       490    526       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       528    564       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       566    601       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       603    639       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       641    676       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       678    714       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       716    751       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       753    789       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       791    827       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       829    867       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       869    905       EGF-LIKE 22.
FT   DOMAIN       907    943       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       945    981       EGF-LIKE 24.
FT   DOMAIN       983   1019       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      1021   1057       EGF-LIKE 26.
FT   DOMAIN      1059   1095       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      1097   1143       EGF-LIKE 28.
FT   DOMAIN      1145   1181       EGF-LIKE 29.
FT   DOMAIN      1183   1219       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      1221   1265       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      1267   1305       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      1307   1346       EGF-LIKE 33.
FT   DOMAIN      1348   1384       EGF-LIKE 34.
FT   DOMAIN      1387   1426       EGF-LIKE 35.
FT   DOMAIN      1449   1462       EGF-LIKE 36.
FT   DOMAIN      1445   1562       CYS-RICH.
FT   REPEAT      1445   1480       3 X LIN/NOTCH REPEATS.
FT   REPEAT      1481   1522       LIN/NOTCH 1.
FT   REPEAT      1523   1562       LIN/NOTCH 3.
FT   DOMAIN      1865   2075       6 X ANK MOTIF REPEATS.
FT   REPEAT      1865   1910       ANK MOTIF 1.
FT   REPEAT      1912   1942       ANK MOTIF 2.
FT   REPEAT      1944   1975       ANK MOTIF 3.
FT   REPEAT      1978   2009       ANK MOTIF 4.
FT   REPEAT      2011   2042       ANK MOTIF 5.
FT   REPEAT      2044   2075       ANK MOTIF 6.
FT   DISULFID      24     37       BY SIMILARITY.
FT   DISULFID      31     46       BY SIMILARITY.
FT   DISULFID      63     74       BY SIMILARITY.
```

```
FT   DISULFID     68     87       BY SIMILARITY.
FT   DISULFID     89     98       BY SIMILARITY.
FT   DISULFID    106    117       BY SIMILARITY.
FT   DISULFID    111    127       BY SIMILARITY.
FT   DISULFID    129    138       BY SIMILARITY.
FT   DISULFID    144    155       BY SIMILARITY.
FT   DISULFID    149    164       BY SIMILARITY.
FT   DISULFID    166    175       BY SIMILARITY.
FT   DISULFID    182    195       BY SIMILARITY.
FT   DISULFID    189    204       BY SIMILARITY.
FT   DISULFID    206    215       BY SIMILARITY.
FT   DISULFID    222    233       BY SIMILARITY.
FT   DISULFID    227    243       BY SIMILARITY.
FT   DISULFID    245    254       BY SIMILARITY.
FT   DISULFID    261    272       BY SIMILARITY.
FT   DISULFID    266    281       BY SIMILARITY.
FT   DISULFID    283    292       BY SIMILARITY.
FT   DISULFID    299    312       BY SIMILARITY.
FT   DISULFID    306    321       BY SIMILARITY.
FT   DISULFID    323    332       BY SIMILARITY.
FT   DISULFID    339    350       BY SIMILARITY.
FT   DISULFID    344    359       BY SIMILARITY.
FT   DISULFID    361    370       BY SIMILARITY.
FT   DISULFID    376    387       BY SIMILARITY.
FT   DISULFID    381    398       BY SIMILARITY.
FT   DISULFID    400    409       BY SIMILARITY.
FT   DISULFID    416    429       BY SIMILARITY.
FT   DISULFID    423    438       BY SIMILARITY.
FT   DISULFID    440    449       BY SIMILARITY.
FT   DISULFID    456    467       BY SIMILARITY.
FT   DISULFID    461    476       BY SIMILARITY.
FT   DISULFID    478    487       BY SIMILARITY.
FT   DISULFID    494    505       BY SIMILARITY.
FT   DISULFID    499    514       BY SIMILARITY.
FT   DISULFID    516    525       BY SIMILARITY.
FT   DISULFID    532    543       BY SIMILARITY.
FT   DISULFID    537    552       BY SIMILARITY.
FT   DISULFID    554    563       BY SIMILARITY.
FT   DISULFID    570    580       BY SIMILARITY.
FT   DISULFID    575    589       BY SIMILARITY.
FT   DISULFID    591    600       BY SIMILARITY.
FT   DISULFID    607    618       BY SIMILARITY.
FT   DISULFID    612    627       BY SIMILARITY.
FT   DISULFID    629    638       BY SIMILARITY.
FT   DISULFID    645    655       BY SIMILARITY.
FT   DISULFID    650    664       BY SIMILARITY.
FT   DISULFID    666    675       BY SIMILARITY.
FT   DISULFID    682    693       BY SIMILARITY.
FT   DISULFID    687    702       BY SIMILARITY.
FT   DISULFID    704    713       BY SIMILARITY.
FT   DISULFID    720    730       BY SIMILARITY.
FT   DISULFID    725    739       BY SIMILARITY.
FT   DISULFID    741    750       BY SIMILARITY.
FT   DISULFID    757    768       BY SIMILARITY.
FT   DISULFID    762    777       BY SIMILARITY.
FT   DISULFID    779    788       BY SIMILARITY.
FT   DISULFID    795    806       BY SIMILARITY.
FT   DISULFID    800    815       BY SIMILARITY.
FT   DISULFID    817    826       BY SIMILARITY.
FT   DISULFID    833    844       BY SIMILARITY.
FT   DISULFID    838    855       BY SIMILARITY.
FT   DISULFID    857    866       BY SIMILARITY.
FT   DISULFID    873    884       BY SIMILARITY.
FT   DISULFID    878    893       BY SIMILARITY.
FT   DISULFID    895    904       BY SIMILARITY.
FT   DISULFID    911    922       BY SIMILARITY.
FT   DISULFID    916    931       BY SIMILARITY.
FT   DISULFID    933    942       BY SIMILARITY.
FT   DISULFID    987    998       BY SIMILARITY.
FT   DISULFID    992   1007       BY SIMILARITY.
FT   DISULFID   1009   1018       BY SIMILARITY.
FT   DISULFID   1025   1036       BY SIMILARITY.
FT   DISULFID   1030   1045       BY SIMILARITY.
```

```
FT   DISULFID   1047   1056       BY SIMILARITY.
FT   DISULFID   1063   1074       BY SIMILARITY.
FT   DISULFID   1068   1083       BY SIMILARITY.
FT   DISULFID   1085   1094       BY SIMILARITY.
FT   DISULFID   1101   1122       BY SIMILARITY.
FT   DISULFID   1116   1131       BY SIMILARITY.
FT   DISULFID   1133   1142       BY SIMILARITY.
FT   DISULFID   1149   1160       BY SIMILARITY.
FT   DISULFID   1154   1169       BY SIMILARITY.
FT   DISULFID   1171   1180       BY SIMILARITY.
FT   DISULFID   1187   1198       BY SIMILARITY.
FT   DISULFID   1192   1207       BY SIMILARITY.
FT   DISULFID   1209   1218       BY SIMILARITY.

Query Match             13.0%;  Score 75.5;  DB 1;  Length 2531;
Best Local Similarity   29.1%;  Pred. No. 9;
Matches   25;  Conservative   4;  Mismatches  24;  Indels  33;  Gaps   5;

Qy    48 GLECQSV---------TSRGDLATCPRGFAVTGCTC--GSACGSWDVR 84
          :||           !:|!  !  |!||  !! ! !| |
Db  1301 GRRCESVINGCRGKPCKNGGVCAVASNTARGFICRCRCPAGF-EGATCENDARTCGSLRCL 1358

Qy    85 AETTC------HCQCAGMDWTGARC 103
          ||            | |  | |
Db  1359 NGGTCISGPRSPTCLCLG-SFTGPEC 1383


RESULT   5
NTC1_HUMAN
ID   NTC1_HUMAN     STANDARD;      PRT;   2444 AA.
AC   P46531;
DT   01-NOV-1995 (Rel. 32, Created).
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLOCATION-
DE   ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).
GN   NOTCH1 OR TAN1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91347367.
RA   Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
RA   Smith S.D., Sklar J.;
RT   "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT   chromosomal translocations in T lymphoblastic neoplasms.";
RL   Cell 66:649-661(1991)
CC   -!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
CC       ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC       IN SOME T-CELL NEOPLASMS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC       BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC       IS FOUND MAINLY IN LYMPHOID TISSUES.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
```

```
DR   EMBL; M73980; AAA60614.1; -.
DR   HSSP; P00740; 1IXA.
DR   MIM; 190198; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 20.
DR   PROSITE; PS00022; EGF_1; 34.
```

```
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 18.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
KW   Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Signal; Glycoprotein.
FT   SIGNAL        1      18       POTENTIAL.
FT   CHAIN        19   >2444       NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1.
FT   DOMAIN       19    1736       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1737    1757       POTENTIAL.
FT   DOMAIN     1758   >2444       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       20      58       EGF-LIKE 1.
FT   DOMAIN       59      99       EGF-LIKE 2.
FT   DOMAIN      102     139       EGF-LIKE 3.
FT   DOMAIN      140     176       EGF-LIKE 4.
FT   DOMAIN      178     216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      218     255       EGF-LIKE 6.
FT   DOMAIN      257     293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      295     333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      335     371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      372     410       EGF-LIKE 10.
FT   DOMAIN      412     450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      452     488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      490     526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      528     564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      566     601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      603     639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      641     676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      678     714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      716     751       EGF-LIKE 19.
FT   DOMAIN      753     789       EGF-LIKE 20.
FT   DOMAIN      791     827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      829     868       EGF-LIKE 22.
FT   DOMAIN      870     906       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      908     944       EGF-LIKE 24.
FT   DOMAIN      946     982       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      984    1020       EGF-LIKE 26.
FT   DOMAIN     1022    1058       EGF-LIKE 27.
FT   DOMAIN     1060    1096       EGF-LIKE 28.
FT   DOMAIN     1098    1144       EGF-LIKE 29.
FT   DOMAIN     1146    1182       EGF-LIKE 30.
FT   DOMAIN     1184    1220       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1222    1266       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1268    1306       EGF-LIKE 33.
FT   DOMAIN     1308    1347       EGF-LIKE 34.
FT   DOMAIN     1349    1385       EGF-LIKE 35.
FT   DOMAIN     1388    1427       EGF-LIKE 36.
FT   REPEAT     1446    1563       3 X LIN/NOTCH REPEATS.
FT   REPEAT     1481    1523       LIN/NOTCH 1.
FT   REPEAT     1482    1563       LIN/NOTCH 2.
FT   REPEAT     1524    1563       LIN/NOTCH 3.
FT   DOMAIN     1876    2087       6 X ANK MOTIF REPEATS.
FT   REPEAT     1921    1954       ANK MOTIF 1.
FT   REPEAT     1923    1987       ANK MOTIF 2.
FT   REPEAT     1956    2021       ANK MOTIF 3.
FT   REPEAT     1990    2054       ANK MOTIF 4.
FT   REPEAT     2023    2087       ANK MOTIF 5.
FT   REPEAT     2056    2087       ANK MOTIF 6.
FT   DOMAIN     1576    1579       POLY-VAL.
FT   DOMAIN     1662    1665       POLY-ARG.
FT   DOMAIN     1729    1732       POLY-PRO.
FT   DOMAIN     1741    1744       POLY-ALA.
FT   DOMAIN     1902    1905       POLY-GLU.
FT   DOMAIN     2260    2263       POLY-GLY.
FT   DOMAIN     2404    2407       POLY-GLN.
FT   DOMAIN     2411    2418       POLY-PRO.
FT   DISULFID     24      37       BY SIMILARITY.
FT   DISULFID     31      46       BY SIMILARITY.
FT   DISULFID     48      57       BY SIMILARITY.
FT   DISULFID     63      74       BY SIMILARITY.
FT   DISULFID     68      87       BY SIMILARITY.
FT   DISULFID     89      98       BY SIMILARITY.
FT   DISULFID    106     117       BY SIMILARITY.
FT   DISULFID    111     127       BY SIMILARITY.
FT   DISULFID    129     138       BY SIMILARITY.
FT   DISULFID    144     155       BY SIMILARITY.
FT   DISULFID    149     164       BY SIMILARITY.
FT   DISULFID    166     175       BY SIMILARITY.
FT   DISULFID    182     195       BY SIMILARITY.
FT   DISULFID    189     204       BY SIMILARITY.
FT   DISULFID    206     215       BY SIMILARITY.
FT   DISULFID    222     233       BY SIMILARITY.
FT   DISULFID    227     243       BY SIMILARITY.
FT   DISULFID    245     254       BY SIMILARITY.
FT   DISULFID    261     272       BY SIMILARITY.
FT   DISULFID    266     281       BY SIMILARITY.
FT   DISULFID    283     292       BY SIMILARITY.
FT   DISULFID    299     312       BY SIMILARITY.
FT   DISULFID    306     321       BY SIMILARITY.
FT   DISULFID    323     332       BY SIMILARITY.
FT   DISULFID    339     350       BY SIMILARITY.
FT   DISULFID    344     359       BY SIMILARITY.
FT   DISULFID    361     370       BY SIMILARITY.
FT   DISULFID    376     387       BY SIMILARITY.
FT   DISULFID    381     398       BY SIMILARITY.
FT   DISULFID    400     409       BY SIMILARITY.
FT   DISULFID    416     429       BY SIMILARITY.
FT   DISULFID    423     438       BY SIMILARITY.
FT   DISULFID    440     449       BY SIMILARITY.
FT   DISULFID    456     467       BY SIMILARITY.
FT   DISULFID    461     476       BY SIMILARITY.
FT   DISULFID    478     487       BY SIMILARITY.
FT   DISULFID    494     505       BY SIMILARITY.
FT   DISULFID    499     514       BY SIMILARITY.
FT   DISULFID    516     525       BY SIMILARITY.
FT   DISULFID    532     543       BY SIMILARITY.
FT   DISULFID    537     552       BY SIMILARITY.
FT   DISULFID    554     563       BY SIMILARITY.
FT   DISULFID    570     580       BY SIMILARITY.
FT   DISULFID    575     589       BY SIMILARITY.
FT   DISULFID    591     600       BY SIMILARITY.
FT   DISULFID    607     618       BY SIMILARITY.
FT   DISULFID    612     627       BY SIMILARITY.
FT   DISULFID    629     638       BY SIMILARITY.
FT   DISULFID    645     655       BY SIMILARITY.
FT   DISULFID    650     664       BY SIMILARITY.
FT   DISULFID    666     675       BY SIMILARITY.
FT   DISULFID    682     693       BY SIMILARITY.
FT   DISULFID    687     702       BY SIMILARITY.
FT   DISULFID    704     713       BY SIMILARITY.
FT   DISULFID    720     730       BY SIMILARITY.
FT   DISULFID    725     739       BY SIMILARITY.
FT   DISULFID    741     750       BY SIMILARITY.
FT   DISULFID    757     768       BY SIMILARITY.
FT   DISULFID    762     777       BY SIMILARITY.
FT   DISULFID    779     788       BY SIMILARITY.
FT   DISULFID    795     806       BY SIMILARITY.
FT   DISULFID    800     815       BY SIMILARITY.
FT   DISULFID    817     826       BY SIMILARITY.
FT   DISULFID    833     844       BY SIMILARITY.
FT   DISULFID    838     855       BY SIMILARITY.
FT   DISULFID    857     867       BY SIMILARITY.
FT   DISULFID    874     885       BY SIMILARITY.
FT   DISULFID    879     894       BY SIMILARITY.
FT   DISULFID    896     905       BY SIMILARITY.
FT   DISULFID    912     923       BY SIMILARITY.
FT   DISULFID    917     932       BY SIMILARITY.
FT   DISULFID    934     943       BY SIMILARITY.
FT   DISULFID    988     999       BY SIMILARITY.
FT   DISULFID    993    1008       BY SIMILARITY.
FT   DISULFID   1010    1019       BY SIMILARITY.
FT   DISULFID   1026    1037       BY SIMILARITY.
FT   DISULFID   1031    1046       BY SIMILARITY.
FT   DISULFID   1048    1057       BY SIMILARITY.
FT   DISULFID   1064    1075       BY SIMILARITY.
```

```
FT   DISULFID   1069   1084              BY SIMILARITY.
FT   DISULFID   1086   1095              BY SIMILARITY.
FT   DISULFID   1102   1123              BY SIMILARITY.
FT   DISULFID   1117   1132              BY SIMILARITY.
FT   DISULFID   1134   1143              BY SIMILARITY.
FT   DISULFID   1150   1161              BY SIMILARITY.
FT   DISULFID   1155   1170              BY SIMILARITY.
FT   DISULFID   1172   1181              BY SIMILARITY.
FT   DISULFID   1188   1199              BY SIMILARITY.
FT   DISULFID   1193   1208              BY SIMILARITY.

Query Match              12.8%;   Score 74.5;  DB 1;  Length 2444;
Best Local Similarity    29.1%;   Pred. No. 11;
Matches   25;  Conservative   4;  Mismatches   24;  Indels   33;  Gaps   5;

Qy    48  GLECQSV-----------------------TSRGDLATCPRGFAVTGCTC---GSACGSWDVR  84
              | |:||                        |||:| ||||  |||    |||||||||
Db  1302  GRRCESVINGCKGKPCKNGGTCAVASNTARGFICKCCPAGF--EGATCENDARTCGSLRCL 1359

Qy    85  AETTC------HCCQAGMDWTGARC 103
              ||      |---|:||
Db  1360  NGGTCISGPRSPTCLCLG-PFTGPEC 1384


RESULT   6
CHH2_BOMMO
ID   CHH2_BOMMO       STANDARD;       PRT;     124 AA.
AC   P05687;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CHORION CLASS HIGH-CYSTEINE HCA PROTEIN 12 PRECURSOR (HC-A.12).
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84272653.
RA   Iatrou K., Tsitilou S.G., Kafatos F.C.;
RT   "DNA sequence transfer between two high-cysteine chorion gene
RT   families in the silkmoth Bombyx mori";
RL   Proc. Natl. Acad. Sci. U.S.A. 81:4452-4456(1984).
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF MANY FROM THE EGGSHELL OF THE
CC       SILK MOTH.
CC   -!- SIMILARITY: MEMBER OF THE ALPHA-BRANCH OF CHORION PROTEIN TO WHICH
CC       BELONG CLASSES A, CA AND HCA.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; K02835; AAA27834.1; -.
DR   HSSP; P10968; 2CWG.
KW   Eggshell; Chorion; Repeat; Multigene family; Signal.
FT   SIGNAL     1     21
FT   CHAIN     22    124         CHORION CLASS HIGH-CYSTEINE HCA PROTEIN
FT                               12.
FT   DOMAIN    22     35         LEFT ARM.
FT   DOMAIN    36     83         CENTRAL DOMAIN.
FT   DOMAIN    84    124         RIGHT ARM (G AND C RICH TANDEM REPEATS).
SQ   SEQUENCE   124 AA;  11639 MW;  9FA97C5556935835 CRC64;

Query Match              12.7%;   Score 74;  DB 1;  Pred. No. 0.82;  Length 124;
Best Local Similarity    28.7%;
Matches   29;  Conservative   7;  Mismatches   31;  Indels   34;  Gaps   8;

Qy    9  LPVLGLLVSSKTLCSMEEAINERIQEVAGSLIFRAISSIGLECQSVTSRGDLATCPR--G  66
```

```
Db   51  LPVCG------------ETLVCGRV-PICGGVCFK----GPACAS---GCVSICGRCCG  89
              |||            |: |:  |:|  :  |:  |:: ||      |:||  |  |
Qy   67  FAVTGC-TCGSACGSWDVRAETTCHCQCAGMDWTGA--RCC  104
                                                  |||
Db   90  CGGCGGCGGCGCGGG--------CGCGGCGGCGGCGCGGRCC 122


RESULT   7
NTC1_RAT
ID   NTC1_RAT         STANDARD;       PRT;    2531 AA.
AC   Q07008;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR.
GN   NOTCH1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SCHWANN CELL;
RX   MEDLINE; 92111383.
RA   Weinmaster G., Roberts V.J., Lemke G.;
RT   "A homolog of Drosophila Notch expressed during mammalian
RT   development.";
RL   Development 113:199-205(1991).
RN   [2]
RP   REVISIONS TO 1652-1653.
RA   Weinmaster G.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER
CC       OF TISSUES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
CC       DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
CC       ADULT.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; X57405; CAA40667.1; -.
DR   HSSP; P00740; 1IXA.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 35.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
KW   Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Signal; Glycoprotein.
FT   SIGNAL     1     18         POTENTIAL.
FT   CHAIN     19   2531         NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN    19   1723         EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM 1724   1746        POTENTIAL.
FT   DOMAIN   1747   2531        CYTOPLASMIC (POTENTIAL).
FT   DOMAIN    20     58         EGF-LIKE 1.
FT   DOMAIN    59     99         EGF-LIKE 2.
FT   DOMAIN   102    139         EGF-LIKE 3.
FT   DOMAIN   140    176         EGF-LIKE 4.
FT   DOMAIN   178    216         EGF-LIKE 5; CALCIUM-BINDING (POTENTIAL).
```

```
FT   DOMAIN      218    255   EGF-LIKE 6.
FT   DOMAIN      257    293   EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      295    333   EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      335    371   EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      372    410   EGF-LIKE 10.
FT   DOMAIN      412    450   EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      452    488   EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      490    526   EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      528    564   EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      566    601   EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      603    639   EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      641    676   EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      678    714   EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      716    751   EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      753    789   EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      791    827   EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      829    867   EGF-LIKE 22.
FT   DOMAIN      869    905   EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      907    943   EGF-LIKE 24.
FT   DOMAIN      945    981   EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      983   1019   EGF-LIKE 26.
FT   DOMAIN     1021   1057   EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1059   1095   EGF-LIKE 28.
FT   DOMAIN     1097   1143   EGF-LIKE 29.
FT   DOMAIN     1145   1181   EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1183   1219   EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1221   1265   EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1267   1305   EGF-LIKE 33.
FT   DOMAIN     1307   1346   EGF-LIKE 34.
FT   DOMAIN     1348   1384   EGF-LIKE 35.
FT   DOMAIN     1387   1426   EGF-LIKE 36.
FT   DOMAIN     1449   1462   CYS-RICH.
FT   DOMAIN     1865   2076   6 X ANK MOTIF REPEATS.
FT   REPEAT     1865   1910   ANK MOTIF 1.
FT   REPEAT     1912   1942   ANK MOTIF 2.
FT   REPEAT     1944   1975   ANK MOTIF 3.
FT   REPEAT     1978   2009   ANK MOTIF 4.
FT   REPEAT     2011   2042   ANK MOTIF 5.
FT   REPEAT     2044   2076   ANK MOTIF 6.
FT   DISULFID     24     37   BY SIMILARITY.
FT   DISULFID     31     46   BY SIMILARITY.
FT   DISULFID     48     57   BY SIMILARITY.
FT   DISULFID     63     74   BY SIMILARITY.
FT   DISULFID     68     87   BY SIMILARITY.
FT   DISULFID     89     98   BY SIMILARITY.
FT   DISULFID    106    117   BY SIMILARITY.
FT   DISULFID    111    127   BY SIMILARITY.
FT   DISULFID    129    138   BY SIMILARITY.
FT   DISULFID    144    155   BY SIMILARITY.
FT   DISULFID    149    164   BY SIMILARITY.
FT   DISULFID    166    175   BY SIMILARITY.
FT   DISULFID    182    195   BY SIMILARITY.
FT   DISULFID    189    204   BY SIMILARITY.
FT   DISULFID    206    215   BY SIMILARITY.
FT   DISULFID    222    233   BY SIMILARITY.
FT   DISULFID    227    243   BY SIMILARITY.
FT   DISULFID    245    254   BY SIMILARITY.
FT   DISULFID    261    272   BY SIMILARITY.
FT   DISULFID    266    281   BY SIMILARITY.
FT   DISULFID    283    292   BY SIMILARITY.
FT   DISULFID    299    312   BY SIMILARITY.
FT   DISULFID    306    321   BY SIMILARITY.
FT   DISULFID    323    332   BY SIMILARITY.
FT   DISULFID    339    350   BY SIMILARITY.
FT   DISULFID    344    359   BY SIMILARITY.
FT   DISULFID    361    370   BY SIMILARITY.
FT   DISULFID    376    387   BY SIMILARITY.
FT   DISULFID    381    398   BY SIMILARITY.
FT   DISULFID    400    409   BY SIMILARITY.
FT   DISULFID    416    429   BY SIMILARITY.
FT   DISULFID    423    438   BY SIMILARITY.
FT   DISULFID    440    449   BY SIMILARITY.
FT   DISULFID    456    467   BY SIMILARITY.
FT   DISULFID    461    476   BY SIMILARITY.
FT   DISULFID    478    487   BY SIMILARITY.
FT   DISULFID    494    505   BY SIMILARITY.
FT   DISULFID    499    514   BY SIMILARITY.
FT   DISULFID    516    525   BY SIMILARITY.
FT   DISULFID    532    543   BY SIMILARITY.
FT   DISULFID    537    552   BY SIMILARITY.
FT   DISULFID    554    563   BY SIMILARITY.
FT   DISULFID    570    580   BY SIMILARITY.
FT   DISULFID    575    589   BY SIMILARITY.
FT   DISULFID    591    600   BY SIMILARITY.
FT   DISULFID    607    618   BY SIMILARITY.
FT   DISULFID    612    627   BY SIMILARITY.
FT   DISULFID    629    638   BY SIMILARITY.
FT   DISULFID    645    655   BY SIMILARITY.
FT   DISULFID    650    664   BY SIMILARITY.
FT   DISULFID    666    675   BY SIMILARITY.
FT   DISULFID    682    693   BY SIMILARITY.
FT   DISULFID    687    702   BY SIMILARITY.
FT   DISULFID    704    713   BY SIMILARITY.
FT   DISULFID    720    730   BY SIMILARITY.
FT   DISULFID    725    739   BY SIMILARITY.
FT   DISULFID    741    750   BY SIMILARITY.
FT   DISULFID    757    768   BY SIMILARITY.
FT   DISULFID    762    777   BY SIMILARITY.
FT   DISULFID    779    788   BY SIMILARITY.
FT   DISULFID    795    806   BY SIMILARITY.
FT   DISULFID    800    815   BY SIMILARITY.
FT   DISULFID    817    826   BY SIMILARITY.
FT   DISULFID    833    844   BY SIMILARITY.
FT   DISULFID    838    855   BY SIMILARITY.
FT   DISULFID    857    866   BY SIMILARITY.
FT   DISULFID    873    884   BY SIMILARITY.
FT   DISULFID    878    893   BY SIMILARITY.
FT   DISULFID    895    904   BY SIMILARITY.
FT   DISULFID    911    922   BY SIMILARITY.
FT   DISULFID    916    931   BY SIMILARITY.
FT   DISULFID    933    942   BY SIMILARITY.
FT   DISULFID    987    998   BY SIMILARITY.
FT   DISULFID    992   1007   BY SIMILARITY.
FT   DISULFID   1009   1018   BY SIMILARITY.
FT   DISULFID   1025   1036   BY SIMILARITY.
FT   DISULFID   1030   1045   BY SIMILARITY.
FT   DISULFID   1047   1056   BY SIMILARITY.
FT   DISULFID   1063   1074   BY SIMILARITY.
FT   DISULFID   1068   1083   BY SIMILARITY.
FT   DISULFID   1085   1094   BY SIMILARITY.
FT   DISULFID   1101   1122   BY SIMILARITY.
FT   DISULFID   1116   1131   BY SIMILARITY.
FT   DISULFID   1133   1142   BY SIMILARITY.
FT   DISULFID   1149   1160   BY SIMILARITY.
FT   DISULFID   1154   1169   BY SIMILARITY.
FT   DISULFID   1171   1180   BY SIMILARITY.
FT   DISULFID   1187   1198   BY SIMILARITY.
FT   DISULFID   1192   1207   BY SIMILARITY.
FT   DISULFID   1209   1218   BY SIMILARITY.
FT   DISULFID   1225   1244   BY SIMILARITY.
FT   DISULFID   1238   1253   BY SIMILARITY.
FT   DISULFID   1255   1264   BY SIMILARITY.
FT   DISULFID   1271   1284   BY SIMILARITY.
FT   DISULFID   1276   1293   BY SIMILARITY.
FT   DISULFID   1295   1304   BY SIMILARITY.
FT   DISULFID   1311   1322   BY SIMILARITY.
FT   DISULFID   1316   1334   BY SIMILARITY.
```

```
QueryMatch             Score 74;  DB 1;  Length 2531;
Best Local Similarity 12.7%;       Pred. No. 13;
Matches   36;  Conservative  12;  Mismatches  52;  Indels  40;  Gaps   8;

Qy     4 LCLLLPVL---GLLVSSKTLCSMEEAINERIQEVAGSLIFRAISS--IGLECQS----- 54
           ||| |||   | :  .|  :   :|  |||  : || | |   :|     ||
Db     9 LCITLPALAARGLRCSQPS----GTCLNGGRCEVANGTEACVCSGAFVGQRCQDPSPCL 64
```

```
Qy    54 --------VTSRGDL----ATCPRGFAVTGC--TCGSAC------GSWDVRAETT 88
            |  |     :|| :    | :|| ||    | |  |
Db    65 STPCKNAGTCYVDHGGIVDYACSCCPLGFSGPLCLITPLANACLANPCRNGGTCDLLTLTE 124

Qy    89 CHCQCAGMDWTGARCCRVQP 108
            |:|                |
Db   125 YKCRCP-PGWSGKSCQADP 143

RESULT 8
PRIA_BORBU
ID   PRIA_BORBU   STANDARD;      PRT;    660 AA.
AC   Q45032; O51047;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   PRIMOSOMAL PROTEIN N' (REPLICATION FACTOR Y).
GN   PRIA OR BB0014.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HB19;
RA   Boursaux-Eude C.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE: 98065943;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
CC   -!- FUNCTION: RECOGNIZES A SPECIFIC HAIRPIN SEQUENCE ON PHIX SSDNA.
CC       THIS STRUCTURE IS THEN RECOGNIZED AND BOUND BY PROTEINS PRIB AND
CC       PRIC. FORMATION OF THE PRIMOSOME PROCEEDS WITH THE SUBSEQUENT
CC       ACTIONS OF DNAB, DNAC, DNAT AND PRIMASE. PRIA THEN FUNCTIONS AS A
CC       HELICASE WITHIN THE PRIMOSOME (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE PRIA SUBFAMILY OF HELICASES.
CC   --------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------
DR   EMBL; X97449; CAA66080.1; -.
DR   EMBL; AE001115; AAC66393.1; -.
DR   TIGR; BB0014; -.
DR   PFAM; PF00271; helicase_C; 1.
KW   DNA replication; DNA-binding; ATP-binding; Helicase; Primosome;
KW   Zinc-finger.
FT   NP_BIND     158    165    ATP (POTENTIAL).
FT   SITE        256    259    DEEH BOX.
FT   ZN_FING     370    382    C4-TYPE (POTENTIAL).
FT   ZN_FING     397    413    C4-TYPE (POTENTIAL).
FT   CONFLICT    126    126    P -> L (IN REF. 1).
FT   CONFLICT    555    555    D -> N (IN REF. 1).
SQ   SEQUENCE  660 AA;  77551 MW;  7F1AB19E04F0EFC3 CRC64;

Query Match             12.7%;  Score 73.5;  DB 1;  Length 660;
Best Local Similarity   29.2%;  Pred. No. 4.2;  -.
Matches   28;  Conservative  13;  Mismatches  26;  Indels  29;  Gaps  7;
```

```
Qy    16 VSSKTLCSMEEAINERIQEVAGSLIFRAISSIG----LECQSVTSRGDLATCPRGFAVTG 71
            :|| :  |:::::||:|:       |||:||     : ||  | | ||
Db   337 ISSELLYSIQKSLNEKRQ----SLIF--INKRGYLKNLECNEC---GHIICCP------N 381

Qy    72 CTCG--------SACGSWDVRAETTCHC-QCAGMD 97
            |:|                |  ||  ||:| ||
Db   382 CSFGLIYHKKENKLLCHYCSYKTKTASHCPQCESKD 417

RESULT 9
ADP1_YEAST
ID   ADP1_YEAST    STANDARD;      PRT;   1049 AA.
AC   P25371;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   PROBABLE ATP-DEPENDENT PERMEASE PRECURSOR.
GN   ADP1 OR YCR011C OR YCR11C OR YCR105.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE: 92160395.
RA   Purnelle B., Skala J., Goffeau A.;
RT   "The product of the YCR105 gene located on the chromosome III from
RT   Saccharomyces cerevisiae presents homologies to ATP-dependent
RT   permeases.";
RL   Yeast 7:867-872(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE: 92327849.
RA   Skala J., Purnelle B., Goffeau A.;
RT   "The complete sequence of a 10.8 kb segment distal of SUF2 on the
RT   right arm of chromosome III from Saccharomyces cerevisiae reveals
RT   seven open reading frames including the RVS161, ADP1 and PGK genes.";
RL   Yeast 8:409-417(1992).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). MDR SUBFAMILY.
CC   --------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------
DR   EMBL; X59720; CAA42328.1; -.
DR   PIR; S19421; S19421.
DR   PIR; S40914; S40914.
DR   SGD; L0000049; ADP1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR   PFAM; PF00005; ABC_tran; 1.
KW   ATP-binding; Transmembrane; Glycoprotein; Transport; Signal.
FT   SIGNAL        1     25
FT   CHAIN        26   1049    PROBABLE ATP-DEPENDENT PERMEASE.
FT   NP_BIND     423    430    ATP (BY SIMILARITY).
FT   TRANSMEM    325    345    POTENTIAL.
FT   TRANSMEM    464    481    POTENTIAL.
FT   TRANSMEM    794    814    POTENTIAL.
FT   TRANSMEM    829    849    POTENTIAL.
FT   TRANSMEM    878    898    POTENTIAL.
FT   TRANSMEM    910    930    POTENTIAL.
FT   TRANSMEM    938    958    POTENTIAL.
FT   TRANSMEM   1001   1021    POTENTIAL.
FT   TRANSMEM   1025   1045    POTENTIAL.
FT   CARBOHYD     50     50    POTENTIAL.
FT   CARBOHYD    114    114    POTENTIAL.
FT   CARBOHYD    165    165    POTENTIAL.
FT   CARBOHYD    221    221    POTENTIAL.
```

```
FT   CARBOHYD     815    815               POTENTIAL.
FT   CARBOHYD     935    935               POTENTIAL.
FT   CARBOHYD     960    960               POTENTIAL.
FT   CARBOHYD     971    971               POTENTIAL.
SQ   SEQUENCE     1049 AA;   117231 MW;  ABC9CE54BCFDF6A3 CRC64;

Query Match             12.6%;   Score 73;  DB 1;  Length 1049;
Best Local Similarity   35.9%;   Pred. No. 7.1;
Matches    23;  Conservative   4;  Mismatches   21;  Indels   16;  Gaps   6;

Qy      50  ECQSVTSRGDLATCPRGFAVTGCT---CG----SACGSWD--VRAET-TCHCQCAGMDWT  99
                |: |||  :   |  |:||||||  |:|:||  ||||||  ||   | |||| || |
Db      71  ECNSYTGCR---CECIEGFAGDDCSLPLCGGLSPDESGNKDRPIRAQNDTCHCD---NGWG  124

Qy     100  GARC  103
                |  |
Db     125  GINC  128


RESULT  10
BAR3_CHITE
ID   BAR3_CHITE     STANDARD;      PRT;  1700 AA.
AC   Q03376;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   BALBIANI RING PROTEIN 3 PRECURSOR.
GN   BR3.
OS   Chironomus tentans (Midge).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC   Chironomoidea; Chironomidae; Chironominae; Chironomus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SALIVARY GLAND;
RX   MEDLINE; 90172404.
RA   Paulsson G., Lendahl U., Galli J., Ericsson C., Wieslander L.;
RT   "The Balbiani ring 3 gene in Chironomus tentans has a diverged
RT   repetitive structure split by many introns.";
RL   J. Mol. Biol. 211:331-349(1990).
CC   -!- FUNCTION: USED BY THE LARVAE TO CONSTRUCT A SUPRAMOLECULAR
CC       STRUCTURE, THE LARVAL TUBE. BALBIANI RING PROTEIN 3 COULD PLAY A
CC       ROLE AS A TRANSPORT PROTEIN THAT BINDS TO OTHER PROTEINS
CC       INTRACELLULARLY AND IN THE GLAND LUMEN IN ORDER TO PREVENT THESE
CC       FROM FORMING WATER-INSOLUBLE FIBERS TOO EARLY.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: SALIVARY GLAND.
CC   -!- DOMAIN: HAS 82 APPROXIMATE REPEATS OF CYS-X-CYS-X-CYS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use  by  non-profit institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X52263; CAA36506.1; -.
DR   PIR; S08167; S08167.
DR   HSSP; P18055; 2MRB.
DR   PRINTS; PR00876; MTNEMATODE.
KW   Repeat; Signal.
FT   SIGNAL       1       ?                 POTENTIAL.
FT   CHAIN        ?      1700               BALBIANI RING PROTEIN 3.
SQ   SEQUENCE     1700 AA;   186145 MW;  3420ZB28521B0815 CRC64;

Query Match             12.6%;   Score 73;  DB 1;  Length 1700;
Best Local Similarity   32.6%;   Pred. No. 11;
Matches    14;  Conservative   7;  Mismatches   12;  Indels   10;  Gaps   2;

Qy      62  TCPRGFAV---TGCTC------GSACGSWDVRAETTCHCQCA  94
```

```
Db    1508  |:||::|         ||::||.:|||       ||:  :::||| |::|
            TCKQGFSPSKSGCKCILECNKKDPGCGAKKIWCQETCKCECA  1550


RESULT  11
NOTC_BRARE
ID   NOTC_BRARE     STANDARD;      PRT;  2437 AA.
AC   P46530;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN PRECURSOR.
GN   NOTCH.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC   Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC   Cyprinoidea; Cyprinidae; Rasborinae; Danio.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 94128602.
RA   Bierkamp C., Campos-Ortega J.A.;
RT   "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RT   its pattern of transcription during early embryogenesis.";
RL   Mech. Dev. 43:87-100(1993).
CC   -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
CC       EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
CC       NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN ALL CELLS IN PREGASTRULATION
CC       STAGES. DURING GASTRULATION IS DIFFERENTIALLY EXPRESSED,
CC       ACCUMULATING PREDOMINANTLY IN THE PRECHORDAL MESODERM AND
CC       NOTOCHORD. AT THE END OF GASTRULATION, EXPRESSED ALONG THE
CC       ANTERIOR-POSTERIOR AXIS INCLUDING THE DEVELOPING NEURAL PLATE
CC       AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING
CC       BRAIN AND HEAD REGIONS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use  by  non-profit institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X69088; CAA48831.1; -.
DR   HSSP; P00740; 1IXA.
DR   ZFIN; ZDB-GENE-990415-173; NOTCH.
DR   PRINTS; PR00009; EGFTGF.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PROSITE; PS00010; ASX_HYDROXYL; 23.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 28.
DR   PROSITE; PS01187; EGF_CA; 22.
DR   PFAM; PF00008; EGF; 36.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
KW   Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Signal; Glycoprotein.
FT   SIGNAL       1      20                 POTENTIAL.
FT   CHAIN       21    2437                 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN.
FT   DOMAIN      21    1724                 EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1725    1747                 POTENTIAL.
FT   DOMAIN    1748    2437                 CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      21      57                 EGF-LIKE 1.
FT   DOMAIN      58      98                 EGF-LIKE 2.
FT   DOMAIN     101     138                 EGF-LIKE 3.
FT   DOMAIN     139     175                 EGF-LIKE 4.
FT   DOMAIN     177     215                 EGF-LIKE 5. CALCIUM-BINDING (POTENTIAL).
```

```
FT   DOMAIN       217    254  EGF-LIKE 6.
FT   DOMAIN       256    292  EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       294    332  EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       334    370  EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       371    409  EGF-LIKE 10.
FT   DOMAIN       411    449  EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       451    487  EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       489    524  EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       526    562  EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       564    599  EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       601    637  EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       639    674  EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       676    712  EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       714    749  EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       751    787  EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       789    825  EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       827    865  EGF-LIKE 22.
FT   DOMAIN       867    903  EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       905    941  EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       943    979  EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       981   1017  EGF-LIKE 26.
FT   DOMAIN      1019   1055  EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      1057   1093  EGF-LIKE 28.
FT   DOMAIN      1095   1141  EGF-LIKE 29.
FT   DOMAIN      1143   1179  EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      1181   1217  EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      1219   1263  EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      1265   1303  EGF-LIKE 33.
FT   DOMAIN      1305   1344  EGF-LIKE 34.
FT   DOMAIN      1346   1382  EGF-LIKE 35.
FT   DOMAIN      1385   1423  EGF-LIKE 36.
FT   REPEAT      1446   1561  3 X LIN/NOTCH REPEATS.
FT   REPEAT      1446   1486  LIN/NOTCH 1.
FT   REPEAT      1487   1520  LIN/NOTCH 2.
FT   REPEAT      1521   1561  LIN/NOTCH 3.
FT   DOMAIN      1861   2074  6 X ANK MOTIF REPEATS.
FT   REPEAT      1861   1891  ANK MOTIF 1.
FT   REPEAT      1892   1940  ANK MOTIF 1.
FT   REPEAT      1941   1974  ANK MOTIF 1.
FT   REPEAT      1975   2007  ANK MOTIF 1.
FT   REPEAT      2008   2040  ANK MOTIF 1.
FT   REPEAT      2041   2074  ANK MOTIF 1.
FT   DOMAIN      2265   2276  POLY-GLN (OPA-REPEAT).
FT   DISULFID      25     35  BY SIMILARITY.
FT   DISULFID      29     45  BY SIMILARITY.
FT   DISULFID      47     56  BY SIMILARITY.
FT   DISULFID      62     73  BY SIMILARITY.
FT   DISULFID      67     86  BY SIMILARITY.
FT   DISULFID      88     97  BY SIMILARITY.
FT   DISULFID     105    116  BY SIMILARITY.
FT   DISULFID     110    126  BY SIMILARITY.
FT   DISULFID     128    137  BY SIMILARITY.
FT   DISULFID     143    154  BY SIMILARITY.
FT   DISULFID     148    163  BY SIMILARITY.
FT   DISULFID     165    174  BY SIMILARITY.
FT   DISULFID     181    194  BY SIMILARITY.
FT   DISULFID     188    203  BY SIMILARITY.
FT   DISULFID     205    214  BY SIMILARITY.
FT   DISULFID     221    232  BY SIMILARITY.
FT   DISULFID     226    242  BY SIMILARITY.
FT   DISULFID     244    253  BY SIMILARITY.
FT   DISULFID     260    271  BY SIMILARITY.
FT   DISULFID     265    280  BY SIMILARITY.
FT   DISULFID     282    291  BY SIMILARITY.
FT   DISULFID     298    311  BY SIMILARITY.
FT   DISULFID     305    320  BY SIMILARITY.
FT   DISULFID     322    331  BY SIMILARITY.
FT   DISULFID     338    349  BY SIMILARITY.
FT   DISULFID     343    358  BY SIMILARITY.
FT   DISULFID     360    369  BY SIMILARITY.
FT   DISULFID     375    386  BY SIMILARITY.
FT   DISULFID     380    397  BY SIMILARITY.
FT   DISULFID     399    408  BY SIMILARITY.

FT   DISULFID     415    428  BY SIMILARITY.
FT   DISULFID     422    437  BY SIMILARITY.
FT   DISULFID     439    448  BY SIMILARITY.
FT   DISULFID     455    466  BY SIMILARITY.
FT   DISULFID     460    475  BY SIMILARITY.
FT   DISULFID     477    486  BY SIMILARITY.
FT   DISULFID     493    503  BY SIMILARITY.
FT   DISULFID     498    512  BY SIMILARITY.
FT   DISULFID     514    523  BY SIMILARITY.
FT   DISULFID     530    541  BY SIMILARITY.
FT   DISULFID     535    550  BY SIMILARITY.
FT   DISULFID     552    561  BY SIMILARITY.
FT   DISULFID     568    578  BY SIMILARITY.
FT   DISULFID     573    587  BY SIMILARITY.
FT   DISULFID     589    598  BY SIMILARITY.
FT   DISULFID     605    616  BY SIMILARITY.
FT   DISULFID     610    625  BY SIMILARITY.
FT   DISULFID     627    636  BY SIMILARITY.
FT   DISULFID     643    653  BY SIMILARITY.
FT   DISULFID     648    662  BY SIMILARITY.
FT   DISULFID     664    673  BY SIMILARITY.
FT   DISULFID     680    691  BY SIMILARITY.
FT   DISULFID     685    700  BY SIMILARITY.
FT   DISULFID     702    711  BY SIMILARITY.
FT   DISULFID     718    728  BY SIMILARITY.
FT   DISULFID     723    737  BY SIMILARITY.
FT   DISULFID     739    748  BY SIMILARITY.
FT   DISULFID     755    766  BY SIMILARITY.
FT   DISULFID     760    775  BY SIMILARITY.
FT   DISULFID     777    786  BY SIMILARITY.
FT   DISULFID     793    804  BY SIMILARITY.
FT   DISULFID     798    813  BY SIMILARITY.
FT   DISULFID     815    824  BY SIMILARITY.
FT   DISULFID     831    842  BY SIMILARITY.
FT   DISULFID     836    853  BY SIMILARITY.
FT   DISULFID     855    864  BY SIMILARITY.
FT   DISULFID     871    882  BY SIMILARITY.
FT   DISULFID     876    891  BY SIMILARITY.
FT   DISULFID     893    902  BY SIMILARITY.
FT   DISULFID     909    920  BY SIMILARITY.
FT   DISULFID     914    929  BY SIMILARITY.
FT   DISULFID     931    940  BY SIMILARITY.
FT   DISULFID     947    958  BY SIMILARITY.
FT   DISULFID     952    967  BY SIMILARITY.
FT   DISULFID     969    978  BY SIMILARITY.
FT   DISULFID    1023   1034  BY SIMILARITY.
FT   DISULFID    1028   1043  BY SIMILARITY.
FT   DISULFID    1045   1054  BY SIMILARITY.
FT   DISULFID    1061   1072  BY SIMILARITY.
FT   DISULFID    1066   1081  BY SIMILARITY.
FT   DISULFID    1083   1092  BY SIMILARITY.
FT   DISULFID    1099   1120  BY SIMILARITY.
FT   DISULFID    1114   1129  BY SIMILARITY.
FT   DISULFID    1131   1140  BY SIMILARITY.
FT   DISULFID    1147   1158  BY SIMILARITY.
FT   DISULFID    1152   1167  BY SIMILARITY.
FT   DISULFID    1169   1178  BY SIMILARITY.
FT   DISULFID    1185   1196  BY SIMILARITY.
FT   DISULFID    1190   1205  BY SIMILARITY.
FT   DISULFID    1207   1216  BY SIMILARITY.
```

```
Query Match              12.6%;   Score  73;  DB 1;  Length 2437;
Best Local Similarity    30.6%;   Pred. No. 15;
  Matches  22;  Conservative   3;  Mismatches  21;  Indels  26;  Gaps   5;

Qy     51 CQSVTSRGD-----LATCPRGFAVTGCTC--------GSACGSWDVRAETTCHC 91
          ||:   |       ||  :  |:|   |            :  |:   | ||
Db   1028 CQNGGSCQDGYGTYKCTCPHGY--TGLNCQSLVRWCDSSPCKNGGSC--WQQGASFTCQC 1083

Qy     92 QCAGMDWTGARC 103
          || |   ||
Db   1084 ---ASGWTGIYC 1092
```

```
RESULT 12
LMG2_HUMAN
ID   LMG2_HUMAN     STANDARD;      PRT;  1193 AA.
AC   Q13753; Q13752; Q14941; Q02535; Q02537;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   LAMININ GAMMA-2 CHAIN PRECURSOR.
GN   LAMC2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 96230326.
RA   Airenne T., Haakana H., Sainio K., Kallunki P., Kallunki P.,
RA   Sariola H., Tryggvason K.;
RT   "Structure of the human laminin gamma 2 chain gene (LAMC2):
RT   alternative splicing with different tissue distribution of two
RT   transcripts.";
RL   Genomics 32:54-64(1996).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1090-1114.
RC   TISSUE=EPIDERMIS, AND KERATINOCYTES;
RX   MEDLINE; 94139694.
RA   Vailly J., Verrando P., Champliaud M.F., Gerecke D., Wagman D.W.,
RA   Baudoin C., Aberdam D., Burgeson R., Bauer E., Ortonne J.P.;
RT   "The 100-kDa chain of nicein/kalinin is a laminin B2 chain variant.";
RL   Eur. J. Biochem. 219:209-218(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FIBROSARCOMA;
RX   MEDLINE; 93016279.
RA   Kallunki P., Sainio K., Eddy R., Byers M., Kallunki T.,
RA   Sariola H., Beck K., Hirvonen H., Shows T.B., Tryggvason K.;
RT   "A truncated laminin chain homologous to the B2 chain: structure,
RT   spatial expression, and chromosomal assignment.";
RL   J. Cell Biol. 119:679-693(1992).
CC   -!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
CC       IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CC       CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC       WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC   -!- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
CC       DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
CC       TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
CC       COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC       THE GAMMA-2 CHAIN IS A SUBUNIT OF LAMININ-5 (EPILIGRIN/KALININ/
CC       NICEIN).
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC       MEMBRANES (MAJOR COMPONENT).
CC   -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: THE LARGE VARIANT IS EXPRESSED ONLY IN
CC       SPECIFIC EPITHELIAL CELLS OF EMBRYONIC AND NEONATAL TISSUES. IN
CC       17-WEEK OLD EMBRYO THE SMALL VARIANT IS FOUND IN CEREBRAL CORTEX,
CC       LUNG, AND DISTAL TUBES OF KIDNEY, BUT NOT IN EPITHELIA EXCEPT FOR
CC       DISTAL TUBULI.
CC   -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC       WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC   -!- DOMAIN: DOMAIN IV IS GLOBULAR.
CC   -!- DISEASE: JUNCTIONAL EPIDERMOLYSIS BULLOSA GRAVIS IS A BLISTERING
CC       DISORDER IN SKIN THAT IS CHARACTERIZED BY A SEPARATION OF BASAL
CC       CELLS FROM THE BASEMENT MEMBRANE DUE TO A DECREASED NUMBER OF
CC       HEMIDESMOSOMES. LAMININ-5 IS MISSING FROM THE BASEMENT MEMBRANE OF
CC       PATIENTS WITH THE GRAVIS FORM OF EPIDERMOLYSIS BULLOSA.
CC   -!- SIMILARITY: CONTAINS 7.5 LAMININ EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
```

```
DR   EMBL; U31201; AAC50457.1; -.
DR   EMBL; U31178; AAC50457.1; JOINED.
DR   EMBL; U31179; AAC50457.1; JOINED.
DR   EMBL; U31180; AAC50457.1; JOINED.
DR   EMBL; U31181; AAC50457.1; JOINED.
DR   EMBL; U31182; AAC50457.1; JOINED.
DR   EMBL; U31183; AAC50457.1; JOINED.
DR   EMBL; U31184; AAC50457.1; JOINED.
DR   EMBL; U31186; AAC50457.1; JOINED.
DR   EMBL; U31187; AAC50457.1; JOINED.
DR   EMBL; U31188; AAC50457.1; JOINED.
DR   EMBL; U31189; AAC50457.1; JOINED.
DR   EMBL; U31190; AAC50457.1; JOINED.
DR   EMBL; U31191; AAC50457.1; JOINED.
DR   EMBL; U31192; AAC50457.1; JOINED.
DR   EMBL; U31193; AAC50457.1; JOINED.
DR   EMBL; U31194; AAC50457.1; JOINED.
DR   EMBL; U31195; AAC50457.1; JOINED.
DR   EMBL; U31196; AAC50457.1; JOINED.
DR   EMBL; U31197; AAC50457.1; JOINED.
DR   EMBL; U31198; AAC50457.1; JOINED.
DR   EMBL; U31199; AAC50457.1; JOINED.
DR   EMBL; U31200; AAC50456.1; -.
DR   EMBL; U31178; AAC50456.1; JOINED.
DR   EMBL; U31179; AAC50456.1; JOINED.
DR   EMBL; U31180; AAC50456.1; JOINED.
DR   EMBL; U31181; AAC50456.1; JOINED.
DR   EMBL; U31182; AAC50456.1; JOINED.
DR   EMBL; U31183; AAC50456.1; JOINED.
DR   EMBL; U31184; AAC50456.1; JOINED.
DR   EMBL; U31186; AAC50456.1; JOINED.
DR   EMBL; U31187; AAC50456.1; JOINED.
DR   EMBL; U31188; AAC50456.1; JOINED.
DR   EMBL; U31189; AAC50456.1; JOINED.
DR   EMBL; U31190; AAC50456.1; JOINED.
DR   EMBL; U31191; AAC50456.1; JOINED.
DR   EMBL; U31192; AAC50456.1; JOINED.
DR   EMBL; U31193; AAC50456.1; JOINED.
DR   EMBL; U31194; AAC50456.1; JOINED.
DR   EMBL; U31195; AAC50456.1; JOINED.
DR   EMBL; U31196; AAC50456.1; JOINED.
DR   EMBL; U31197; AAC50456.1; JOINED.
DR   EMBL; U31198; AAC50456.1; JOINED.
DR   EMBL; X73902; CAA52108.1; -.
DR   EMBL; Z15008; CAA78728.1; -.
DR   EMBL; Z15009; CAA78729.1; -.
DR   HSSP; P02468; 1TLE.
DR   MIM; 150292; -.
DR   PROSITE; PS00022; EGF_1; 4.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 6.
DR   PFAM; PF00052; laminin_B; 1.
DR   PFAM; PF00053; laminin_EGF; 6.
KW   Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW   Laminin EGF-like domain; Cell adhesion; Repeat; Signal;
KW   Alternative splicing.
FT   SIGNAL        1     21
FT   CHAIN        22   1193       LAMININ GAMMA-2 CHAIN.
FT   DOMAIN       28    196       3.5 X LAMININ EGF-LIKE REPEATS (DOMAIN
FT                                V).
FT   DOMAIN       28     83       LAMININ EGF-LIKE 1.
FT   DOMAIN       84    130       LAMININ EGF-LIKE 2.
FT   DOMAIN      139    186       LAMININ EGF-LIKE 3.
FT   DOMAIN      187    196       LAMININ EGF-LIKE 4 (N-TERMINAL).
FT   DOMAIN      197    381       LAMININ DOMAIN IV.
FT   DOMAIN      382    602       4 X LAMININ EGF-LIKE REPEATS (DOMAIN V).
FT   DOMAIN      382    415       LAMININ EGF-LIKE 4 (C-TERMINAL).
FT   DOMAIN      416    461       LAMININ EGF-LIKE 5.
```

```
FT   DOMAIN       462    516       LAMININ EGF-LIKE 6.
FT   DOMAIN       517    572       LAMININ EGF-LIKE 7.
FT   DOMAIN       573    602       LAMININ EGF-LIKE 8 (INCOMPLETE).
FT   DOMAIN       603    611       DOMAIN II AND I.
FT   DOMAIN       611    718       COILED COIL (POTENTIAL).
FT   DOMAIN       811   1076       COILED COIL (POTENTIAL).
FT   DOMAIN      1117   1193       COILED COIL (POTENTIAL).
FT   DISULFID      28     37       BY SIMILARITY.
FT   DISULFID      30     53       BY SIMILARITY.
FT   DISULFID      56     65       BY SIMILARITY.
FT   DISULFID      68     81       BY SIMILARITY.
FT   DISULFID      84     96       BY SIMILARITY.
FT   DISULFID      86    102       BY SIMILARITY.
FT   DISULFID     104    113       BY SIMILARITY.
FT   DISULFID     116    128       BY SIMILARITY.
FT   DISULFID     139    150       BY SIMILARITY.
FT   DISULFID     141    155       BY SIMILARITY.
FT   DISULFID     157    166       BY SIMILARITY.
FT   DISULFID     169    184       BY SIMILARITY.
FT   DISULFID     462    470       BY SIMILARITY.
FT   DISULFID     464    481       BY SIMILARITY.
FT   DISULFID     484    493       BY SIMILARITY.
FT   DISULFID     496    514       BY SIMILARITY.
FT   DISULFID     517    531       BY SIMILARITY.
FT   DISULFID     519    538       BY SIMILARITY.
FT   DISULFID     541    550       BY SIMILARITY.
FT   DISULFID     553    570       BY SIMILARITY.
FT   DISULFID     609    609       INTERCHAIN (PROBABLE).
FT   DISULFID     462    612       INTERCHAIN (PROBABLE).
FT   DISULFID    1184   1184       INTERCHAIN (PROBABLE).
FT   CARBOHYD     342    342       POTENTIAL.
FT   CARBOHYD     362    362       POTENTIAL.
FT   CARBOHYD     942    942       POTENTIAL.
FT   CARBOHYD    1033   1033       POTENTIAL.
FT   VARSPLIC    1110   1193       DQPLSVDEGLVLLEQKLSRAKTQINSQLRPMMSELEERAR
FT                                 QQRGHLHLLETSIDGILADVKNLENIRDNLPPGCYNTQALE
FT                                 QQ -> GM (IN SMALL ISOFORM).
FT   CONFLICT      12     12       F -> L (IN REF. 2).
FT   CONFLICT     473    473       M -> I (IN REF. 3).
FT   CONFLICT     521    521       N -> S (IN REF. 3).
FT   CONFLICT     857    857       R -> P (IN REF. 3).
FT   CONFLICT     883    883       T -> S (IN REF. 2 AND 3).
SQ   SEQUENCE    1193 AA;  130989 MW;  10BFC39E416F4ECE CRC64;

Query Match             12.4%;  Score 72;  DB 1;  Length 1193;
Best Local Similarity   25.4%;  Pred. No. 10;
Matches   31;  Conservative   14;  Mismatches   47;  Indels   30;  Gaps    8;

Qy      4 LCLLLLPVLGLLVSSKTLC----SMEEAINERIQEVAGSLIFRAI---SSIGLECQSV 54
            ||  ||          : ::      |:   :: :: ||:
Db     10 LCFSLLLPAARATSRREVCDCNGKSRQCIFDRELHRQTGN-GFRCLNCNDNTDGIHCEK- 68

Qy     55 TSRGDLATCPRGF------AVTGCTCGSACGSWDVRAETTCHCQC-AGMDWTGARCCRV 106
            |  |          :| | | | ||| || | :| ||||  |||| :
Db     68 ------CKNGFYRHRERDRCLPCNCNSK-GSLSARCDNSGRCSCKPGV--TGARCDRC 116

Qy    107 QP 108
            |
Db    117 LP 118

RESULT 13
NTC3_MOUSE
ID   NTC3_MOUSE     STANDARD;      PRT;  2318 AA.
AC   Q61982;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   NEUROGENIC LOCUS NOTCH 3 PROTEIN.
GN   NOTCH3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR X SWISS WEBSTER;
RX   MEDLINE; 95001556.
RA   Lardelli M., Dalstrand J., Lendahl U.;
RT   "The novel Notch homologue mouse Notch 3 lacks specific epidermal
RT   growth factor-repeats and is expressed in proliferating
RT   neuroepithelium.";
RL   Mech. Dev. 46:123-136(1994).
CC   -!- FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATIONAL ROLE DURING
CC       VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE
CC       DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.
CC   -!- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.
CC   -!- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.
CC   -!- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X74760; CAA52776.1; -.
DR   HSSP; P00740; 1IXA.
DR   MGD; MGI:99460; NOTCH3.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PROSITE; PS00010; ASX_HYDROXYL; 18.
DR   PROSITE; PS00022; EGF_1; 33.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 17.
DR   PFAM; PF00008; EGF; 34.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00066; notch; 3.
KW   Differentiation; Neurogenesis; Repeat; EGF-like domain; Transmembrane;
KW   Glycoprotein.
FT   DOMAIN        1    1643       EXTRACELLULAR.
FT   TRANSMEM   1644    1664       POTENTIAL.
FT   DOMAIN     1665    2318       CYTOPLASMIC.
FT   DOMAIN       39    1374       34 X EGF-TYPE REPEATS.
FT   DOMAIN     1388    1503       3 X LIN/NOTCH REPEATS.
FT   DOMAIN     1784    1998       6 X CDC10/SWI6 REPEATS.
FT   DOMAIN     2242    2261       PEST.
FT   DOMAIN       39      78       EGF-LIKE 1.
FT   DOMAIN       79     119       EGF-LIKE 2.
FT   DOMAIN      120     157       EGF-LIKE 3.
FT   DOMAIN      159     196       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      198     235       EGF-LIKE 5.
FT   DOMAIN      237     273       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      275     313       EGF-LIKE 7.
FT   DOMAIN      315     351       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      352     390       EGF-LIKE 9.
FT   DOMAIN      392     430       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      432     468       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      470     506       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      508     544       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      546     581       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      583     619       EGF-LIKE 15.
FT   DOMAIN      621     656       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      658     694       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      696     731       EGF-LIKE 18.
FT   DOMAIN      735     771       EGF-LIKE 19.
FT   DOMAIN      772     809       EGF-LIKE 20.
FT   DOMAIN      811     848       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      850     886       EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      888     923       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      925     961       EGF-LIKE 24.
FT   DOMAIN      963     999       EGF-LIKE 25.
FT   DOMAIN     1001    1035       EGF-LIKE 26.
```

```
FT   DOMAIN      1037   1083   EGF-LIKE 27.
FT   DOMAIN      1085   1121   EGF-LIKE 28.
FT   DOMAIN      1123   1159   EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      1161   1204   EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      1206   1245   EGF-LIKE 31.
FT   DOMAIN      1247   1288   EGF-LIKE 32.
FT   DOMAIN      1290   1326   EGF-LIKE 33.
FT   DOMAIN      1336   1374   EGF-LIKE 34.
FT   REPEAT      1388   1428   LIN/NOTCH 1.
FT   REPEAT      1429   1467   LIN/NOTCH 2.
FT   REPEAT      1468   1503   LIN/NOTCH 3.
FT   REPEAT      1784   1816   CDC10/SWI6 1.
FT   REPEAT      1817   1865   CDC10/SWI6 2.
FT   REPEAT      1866   1898   CDC10/SWI6 3.
FT   REPEAT      1899   1932   CDC10/SWI6 4.
FT   REPEAT      1933   1965   CDC10/SWI6 5.
FT   REPEAT      1966   1998   CDC10/SWI6 6.
FT   DISULFID      43     55   BY SIMILARITY.
FT   DISULFID      49     66   BY SIMILARITY.
FT   DISULFID      68     77   BY SIMILARITY.
FT   DISULFID      83     94   BY SIMILARITY.
FT   DISULFID      88    107   BY SIMILARITY.
FT   DISULFID     109    118   BY SIMILARITY.
FT   DISULFID     124    135   BY SIMILARITY.
FT   DISULFID     129    145   BY SIMILARITY.
FT   DISULFID     147    156   BY SIMILARITY.
FT   DISULFID     163    175   BY SIMILARITY.
FT   DISULFID     169    184   BY SIMILARITY.
FT   DISULFID     186    195   BY SIMILARITY.
FT   DISULFID     202    213   BY SIMILARITY.
FT   DISULFID     207    223   BY SIMILARITY.
FT   DISULFID     225    234   BY SIMILARITY.
FT   DISULFID     241    252   BY SIMILARITY.
FT   DISULFID     246    261   BY SIMILARITY.
FT   DISULFID     263    272   BY SIMILARITY.
FT   DISULFID     279    292   BY SIMILARITY.
FT   DISULFID     286    301   BY SIMILARITY.
FT   DISULFID     303    312   BY SIMILARITY.
FT   DISULFID     319    330   BY SIMILARITY.
FT   DISULFID     324    339   BY SIMILARITY.
FT   DISULFID     341    350   BY SIMILARITY.
FT   DISULFID     356    367   BY SIMILARITY.
FT   DISULFID     361    378   BY SIMILARITY.
FT   DISULFID     380    389   BY SIMILARITY.
FT   DISULFID     396    409   BY SIMILARITY.
FT   DISULFID     403    418   BY SIMILARITY.
FT   DISULFID     420    429   BY SIMILARITY.
FT   DISULFID     436    447   BY SIMILARITY.
FT   DISULFID     441    456   BY SIMILARITY.
FT   DISULFID     458    467   BY SIMILARITY.
FT   DISULFID     474    485   BY SIMILARITY.
FT   DISULFID     479    494   BY SIMILARITY.
FT   DISULFID     496    505   BY SIMILARITY.
FT   DISULFID     512    523   BY SIMILARITY.
FT   DISULFID     517    532   BY SIMILARITY.
FT   DISULFID     534    543   BY SIMILARITY.
FT   DISULFID     550    560   BY SIMILARITY.
FT   DISULFID     555    569   BY SIMILARITY.
FT   DISULFID     571    580   BY SIMILARITY.
FT   DISULFID     587    598   BY SIMILARITY.
FT   DISULFID     592    607   BY SIMILARITY.
FT   DISULFID     609    618   BY SIMILARITY.
FT   DISULFID     625    635   BY SIMILARITY.
FT   DISULFID     630    644   BY SIMILARITY.
FT   DISULFID     646    655   BY SIMILARITY.
FT   DISULFID     662    673   BY SIMILARITY.
FT   DISULFID     667    682   BY SIMILARITY.
FT   DISULFID     684    693   BY SIMILARITY.
FT   DISULFID     700    710   BY SIMILARITY.
FT   DISULFID     705    719   BY SIMILARITY.
FT   DISULFID     721    730   BY SIMILARITY.
FT   DISULFID     739    750   BY SIMILARITY.
FT   DISULFID     744    759   BY SIMILARITY.

FT   DISULFID     761    770   BY SIMILARITY.
FT   DISULFID     776    787   BY SIMILARITY.
FT   DISULFID     781    797   BY SIMILARITY.
FT   DISULFID     799    808   BY SIMILARITY.
FT   DISULFID     815    827   BY SIMILARITY.
FT   DISULFID     821    836   BY SIMILARITY.
FT   DISULFID     838    847   BY SIMILARITY.
FT   DISULFID     854    865   BY SIMILARITY.
FT   DISULFID     859    874   BY SIMILARITY.
FT   DISULFID     876    885   BY SIMILARITY.
FT   DISULFID     892    902   BY SIMILARITY.
FT   DISULFID     897    911   BY SIMILARITY.
FT   DISULFID     913    922   BY SIMILARITY.
FT   DISULFID     929    940   BY SIMILARITY.
FT   DISULFID     934    949   BY SIMILARITY.
FT   DISULFID     951    960   BY SIMILARITY.
FT   DISULFID     967    978   BY SIMILARITY.
FT   DISULFID     972    987   BY SIMILARITY.
FT   DISULFID     989    998   BY SIMILARITY.
FT   DISULFID    1005   1016   BY SIMILARITY.
FT   DISULFID    1010   1023   BY SIMILARITY.
FT   DISULFID    1025   1034   BY SIMILARITY.
FT   DISULFID    1041   1062   BY SIMILARITY.
FT   DISULFID    1056   1071   BY SIMILARITY.
FT   DISULFID    1073   1082   BY SIMILARITY.
FT   DISULFID    1089   1100   BY SIMILARITY.
FT   DISULFID    1094   1109   BY SIMILARITY.
FT   DISULFID    1111   1120   BY SIMILARITY.
FT   DISULFID    1127   1138   BY SIMILARITY.
FT   DISULFID    1132   1147   BY SIMILARITY.
FT   DISULFID    1149   1158   BY SIMILARITY.
FT   DISULFID    1165   1183   BY SIMILARITY.
FT   DISULFID    1177   1192   BY SIMILARITY.
FT   DISULFID    1194   1203   BY SIMILARITY.
FT   DISULFID    1210   1223   BY SIMILARITY.
FT   DISULFID    1215   1233   BY SIMILARITY.
FT   DISULFID    1235   1244   BY SIMILARITY.
FT   DISULFID    1251   1262   BY SIMILARITY.
FT   DISULFID    1256   1276   BY SIMILARITY.
FT   DISULFID    1278   1287   BY SIMILARITY.
FT   DISULFID    1294   1305   BY SIMILARITY.
FT   DISULFID    1299   1314   BY SIMILARITY.
FT   DISULFID    1316   1325   BY SIMILARITY.
FT   DISULFID    1340   1351   BY SIMILARITY.
FT   DISULFID    1345   1362   BY SIMILARITY.


Query Match                  12.4%;     Score 72;  DB 1;  Length 2318;
Best Local Similarity   32.08;           Pred. No. 18;
Matches   24;  Conservative    7;  Mismatches    7;  Indels   20;  Gaps    6;

Qy     47 IGLECQSVTSRGDLATCPRGFA------------VTGCTCGSA-C---GS-WDVRAETT 88
          :|:||  |:||  ||||  |:  ||::
Db   1301 VGIPCQQ-TARGPRCACPPGLSGPSCRVSRASPSGATNASCASAPCLHGGSCLPVQSVPF 1359

Qy     89 CHCQCAGMDWTGARC 103
          ||| ||:|:||
Db   1360 FRCVCA-PGWGGPRC 1373


RESULT 14
TYR2_HUMAN
ID   TYR2_HUMAN     STANDARD;      PRT;   519 AA.
AC   P40126;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   DOPACHROME TAUTOMERASE PRECURSOR (EC 5.3.12) (DT) (DCT) (DOPACHROME
DE   DELTA-ISOMERASE) (TYROSINASE-RELATED PROTEIN 2) (TRP-2).
GN   DCT OR TYRP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94198295.
RA   Yokoyama K., Suzuki H., Yasumoto K.I., Tomita Y., Shibahara S.;
RT   "Molecular cloning and functional analysis of a cDNA coding for human
RT   DOPAchrome tautomerase/tyrosinase-related protein-2.";
RL   Biochim. Biophys. Acta 1217:317-321(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 9266170.
RA   Cassady J.L., Sturm R.A.;
RT   "Sequence of the human dopachrome tautomerase-encoding TRP-2 cDNA.";
RL   Gene 143:295-298(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94139684.
RA   Bouchard B., del Marmol V., Jackson I.J., Cherif D., Dubertret L.;
RT   "Molecular characterization of a human tyrosinase-related-protein-2
RT   cDNA. Patterns of expression in melanocytic cells.";
RL   Eur. J. Biochem. 219:127-134(1994).
RN   [4]
RP   SEQUENCE OF 1-98 FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 96079088.
RA   Sturm R.A., O'Sullivan B.J., Box N.F., Smith A.G., Smit S.E.,
RA   Puttick E.R., Parsons P.G., Dunn I.S.;
RT   "Chromosomal structure of the human TYRP1 and TYRP2 loci and
RT   comparison of the tyrosinase-related protein gene family.";
RL   Genomics 29:24-34(1995).
RN   [5]
RP   SEQUENCE OF 1-98 FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 95014579.
RA   Yokoyama K., Yasumoto K.I., Suzuki H., Shibahara S.;
RT   "Cloning of the human DOPAchrome tautomerase/tyrosinase-related
RT   protein 2 gene and identification of two regulatory regions required
RT   for its pigment cell-specific expression.";
RL   J. Biol. Chem. 269:27080-27087(1994).
CC   -!- FUNCTION: INVOLVED IN REGULATING EUMELANIN AND PHAEOMELANIN
CC       LEVELS.
CC   -!- CATALYTIC ACTIVITY: L-DOPACHROME = 5,6-DIHYDROXYINDOLE-2-
CC       CARBOXYLATE.
CC   -!- COFACTOR: BINDS TWO ZINC IONS (BY SIMILARITY).
CC   -!- PATHWAY: MELANIN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. MELANOSOMAL.
CC   -!- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY.
CC
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; D17547; BAA04484.1; -.
DR   EMBL; L18967; AAA20870.1; -.
DR   EMBL; S69231; AAC60627.1; -.
DR   EMBL; L38953; AAC41925.1; -.
DR   EMBL; D28767; BAA05956.1; -.
DR   PIR; S43510; S43510.
DR   MIM; 191275; -.
DR   PRINTS; PR00092; TYROSINASE.
DR   PROSITE; PS00497; TYROSINASE_1; 1.
DR   PROSITE; PS00498; TYROSINASE_2; 1.
DR   PFAM; PF00264; tyrosinase; 1.
KW   Isomerase; Zinc; Glycoprotein; Signal; Transmembrane;
KW   Melanin biosynthesis.
FT   SIGNAL        1     23
FT   CHAIN        24    519       DOPACHROME TAUTOMERASE.
FT   DOMAIN       24    472       LUMENAL, MELANOSOME (POTENTIAL).
FT   TRANSMEM    473    493       POTENTIAL.
FT   DOMAIN      494    519       CYTOPLASMIC (POTENTIAL).
FT   METAL       189    189       ZINC A (BY SIMILARITY).
```

```
FT   METAL       211    211       ZINC A (BY SIMILARITY).
FT   METAL       220    220       ZINC A (BY SIMILARITY).
FT   METAL       369    369       ZINC B (BY SIMILARITY).
FT   METAL       373    373       ZINC B (BY SIMILARITY).
FT   METAL       396    396       POTENTIAL.
FT   CARBOHYD    170    170       POTENTIAL.
FT   CARBOHYD    178    178       POTENTIAL.
FT   CARBOHYD    237    237       POTENTIAL.
FT   CARBOHYD    300    300       POTENTIAL.
FT   CARBOHYD    342    342       POTENTIAL.
FT   CARBOHYD    377    377       POTENTIAL.
SQ   SEQUENCE   519 AA;  59145 MW;  AFDDF2176B002A89 CRC64;

Query Match             12.3%;  Score 71.5;  DB 1;  Length 519;
Best Local Similarity   27.4%;  Pred. No. 5.3;
Matches   26;  Conservative   7;  Mismatches   25;  Indels   37;  Gaps   4;

Qy     51 CQSVTSRGDLATCPRGFAVTGCTCGSACG---SWDVRAET-------------- 88
          :|:|  : |||.: ||   : ||:|||   :|||:|
Db     29 CMTVDSLVNKECCPRLGAESANVCGSQQGRGQCTEVRADTRPWSGPYILRNQDDRELWPR 88

Qy     88 ----TCHC--QCAGMD------WTGARCCRVQP 108
              |||   || |     ||| | : |
Db     89 KFFHRTCKCTGNFAGYNCGDCKFGWTGPNCERKKP 123


RESULT 15
ZAN_PIG
ID   ZAN_PIG    STANDARD;    PRT;   2476 AA.
AC   Q28983;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   ZONADHESIN PRECURSOR.
GN   ZAN.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MEISHAN; TISSUE=TESTIS;
RX   MEDLINE; 96064658.
RA   Hardy D.M., Garbers D.L.;
RT   "A sperm membrane protein that binds in a species-specific manner to
RT   the egg extracellular matrix is homologous to von Willebrand
RT   factor.";
RL   J. Biol. Chem. 270:26025-26028(1995).
CC   -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC       OF THE EGG.
CC   -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: IN TESTIS, PRIMARILY IN HAPLOID SPERMATIDS.
CC       NOT IN LUNG, LIVER, HEART, SPLEEN, BRAIN, KIDNEY, EPIDIDYMIS.
CC   -!- DOMAIN: THE UNIQUE N-TERMINAL DOMAIN AND THE MUCIN-LIKE DOMAINS
CC       ARE MISSING FROM THE ZONADHESIN THAT BINDS TO THE EGG
CC       EXTRACELLULAR MATRIX. PROCESSING MIGHT OCCUR DURING SPERM
CC       MATURATION AND/OR CAPACITATION.
CC   -!- DOMAIN: THE UNIQUE N-TERMINAL DOMAIN PROBABLY MEDIATES SPERM
CC       ADHESION TO THE ZONA PELLUCIDA.
CC   -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC       THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC       SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC   -!- DOMAIN: THE VWFD DOMAINS 1 AND 2 MAY MEDIATE COVALENT
CC       OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC   -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed. Usage  by  and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; U40024; AAC48486.1; -.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS00740; MAM_1; 1.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS50060; MAM_2; 2.
DR   PFAM; PF00094; vwd; 4.
KW   Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW   Repeat.
FT   SIGNAL          1     29       POTENTIAL.
FT   CHAIN          30   2476       ZONADHESIN.
FT   DOMAIN         30   2418       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     2419   2439       POTENTIAL.
FT   DOMAIN       2440   2476       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        144    312       MAM.
FT   DOMAIN        319    687       53 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
FT                                  (MUCIN-LIKE DOMAIN).
FT   DOMAIN        688   2370       4.5 X VWFD REPEATS.
FT   DOMAIN        688    799       VWFD 1 (PARTIAL).
FT   DOMAIN        800   1184       VWFD 2.
FT   DOMAIN       1185   1573       VWFD 3.
FT   DOMAIN       1574   1968       VWFD 4.
FT   DOMAIN       1969   2370       VWFD 5.
FT   DOMAIN       2366   2402       EGF-LIKE.
FT   DISULFID     2370   2381       BY SIMILARITY.
FT   DISULFID     2375   2390       BY SIMILARITY.
FT   DISULFID     2392   2401       BY SIMILARITY.
FT   CARBOHYD      109    109       POTENTIAL.
FT   CARBOHYD      269    269       POTENTIAL.
FT   CARBOHYD      735    735       POTENTIAL.
FT   CARBOHYD      758    758       POTENTIAL.
FT   CARBOHYD      833    833       POTENTIAL.
FT   CARBOHYD     1154   1154       POTENTIAL.
FT   CARBOHYD     1329   1329       POTENTIAL.
FT   CARBOHYD     1448   1448       POTENTIAL.
FT   CARBOHYD     1544   1544       POTENTIAL.
FT   CARBOHYD     1596   1596       POTENTIAL.
FT   CARBOHYD     1654   1654       POTENTIAL.
FT   CARBOHYD     1843   1843       POTENTIAL.
FT   CARBOHYD     1965   1965       POTENTIAL.
FT   CARBOHYD     2122   2122       POTENTIAL.
FT   CARBOHYD     2165   2165       POTENTIAL.
FT   CARBOHYD     2178   2178       POTENTIAL.
FT   CARBOHYD     2329   2329       POTENTIAL.
FT   CARBOHYD     2359   2359       POTENTIAL.
FT   SEQUENCE     2476 AA;  270364 MW;  A13B690375A6548C CRC64;

Query Match              12.3%;  Score 71.5;  DB 1;  Length 2476;
Best Local Similarity    20.5%;  Pred. No. 22;
Matches   26;  Conservative  14;  Mismatches  52;  Indels  35;  Gaps   4;

Qy   17  SSKTLCSMEEAINERIQEVAGSLIFRAISSIGLECQS---VTSRGDLATCPRGFAVTGCT   73
         : :|:  |    :|: :  :| ::::  :|: :  |       |:|| ::||  :!:||
Db  641  TEKTIIPTERTIAPTTPQPSPTLVPTQPAAVVMPSTSATTVTPRTTIASCPPNAHFERCA  700

Qy   74  CGSACGSWDVRAETTC--------------------HCQC-------AGMDWTGARC-  104
         | : :|  |  | |                        |  ||   |||    :| |
Db  701  CPVSCQSPTPNCELFCKPGCVCDPGFLFSGSHCVNASSCDCFYNDNYYKLGTDWFSPNCT  760

Qy  104  --CRVQP  108
           | :|
Db  761  EHCHCRP  767

Search completed: May 26, 2000, 04:53:06
Job time: 6674 sec
```

```
                    GenCore version 4.5
            Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    May 26, 2000, 00:32:50 ; Search time 55.06 Seconds
                                    (without alignments)
                                    135.999 Million cell updates/sec

Title:          US-09-099-898-2
Perfect score:  581
Sequence:       1 MKALCLLLPVLGLLVSSKT.........CHCQCAGMDWTGARCCRVQP 108

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       225878 seqs. 69334122 residues

Total number of hits satisfying chosen parameters:    225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :   SPTREMBL_12:*
             1:  sp_archea:*
             2:  sp_bacteria:*
             3:  sp_fungi:*
             4:  sp_human:*
             5:  sp_invertebrate:*
             6:  sp_mammal:*
             7:  sp_mhc:*
             8:  sp_organelle:*
             9:  sp_phage:*
            10:  sp_plant:*
            11:  sp_rodent:*
            12:  sp_virus:*
            13:  sp_vertebrate:*
            14:  sp_unclassified:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	86	14.8	1059	5	P90884	P90884 caenorhabdi
2	84.5	14.5	1476	13	Q90285	Q90285 carassius a
3	82.5	14.2	762	13	O42373	O42373 brachydanio
4	82.5	14.2	1574	11	O88281	O88281 rattus norv
5	82	14.1	2321	4	Q9Y6L8	Q9Y6L8 homo sapien
6	78	13.4	594	5	Q9Y151	Q9Y151 drosophila
7	78	13.4	1106	5	Q17494	Q17494 caenorhabdi
8	76.5	13.2	589	11	O88671	O88671 rattus norv
9	76	13.1	2447	13	O13149	O13149 fugu rubrip
10	75.5	13.0	153	4	O75095	O75095 homo sapien
11	75.5	13.0	387	11	Q06007	Q06007 mus musculu
12	74.5	12.8	75	5	O96388	O96388 perna virid
13	74.5	12.8	585	11	O35675	O35675 mus musculu
14	74.5	12.8	592	11	O88516	O88516 mus musculu
15	74.5	12.8	1698	5	Q94438	Q94438 chironomus
16	73.5	12.7	289	5	P91237	P91237 caenorhabdi
17	73.5	12.7	830	4	O43701	O43701 homo sapien
18	73.5	12.7	830	4	Q14162	Q14162 homo sapien
19	71.5	12.3	105	5	Q17187	Q17187 bombyx mori
20	71.5	12.3	237	4	O75767	O75767 homo sapien
21	71.5	12.3	378	3	O59925	O59925 penicillium
22	71.5	12.3	1300	11	Q9WTL4	Q9wtl4 mus musculu
23	71.5	12.3	1647	5	O45000	O45000 caenorhabdi
24	71.5	12.3	2704	5	O97458	O97458 drosophila
25	71.5	12.3	2906	11	Q9WUH9	Q9wuh9 rattus norv
26	71	12.2	341	5	Q25032	Q25032 haemonchus
27	71	12.2	752	13	O42374	O42374 brachydanio
28	70.5	12.1	626	10	O49438	O49438 arabidopsis
29	70	12.0	127	13	Q9YGK3	Q9ygk3 anguilla ja
30	70	12.0	582	2	Q92892	Q9z892 chlanydia p
31	70	12.0	709	5	O97444	O97444 giardia lam
32	70	12.0	1095	11	Q60784	Q60784 mus musculu
33	70	12.0	3857	11	O88840	O88840 mus musculu
34	69.5	12.0	326	10	P93680	P93680 persea amer
35	69.5	12.0	346	6	O08126	O08126 ovis aries
36	69.5	12.0	1704	5	Q94446	Q94446 chironomus
37	69.5	12.0	2796	4	O95071	O95071 homo sapien
38	69	11.9	266	5	Q94700	Q94700 paramecium
39	69	11.9	379	11	O35883	O35883 rattus norv
40	69	11.9	447	11	Q63348	Q63348 rattus norv
41	69	11.9	625	10	O65428	O65428 arabidopsis
42	69	11.9	2155	4	O75443	O75443 homo sapien
43	68.5	11.8	258	5	O44179	O44179 caenorhabdi
44	68.5	11.8	377	5	O61699	O61699 branchiosto
45	68.5	11.8	379	11	Q9WUA1	Q9wual mus musculu

ALIGNMENTS

```
RESULT  1
P90884
ID   P90884        PRELIMINARY;       PRT;   1059 AA.
AC   P90884;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   F53B6.2 PROTEIN.
GN   F53B6.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   WHITE S.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z81086; CAB01121.1; -.
DR   PFAM; PF00047; ig; 1.
SQ   SEQUENCE   1059 AA;  117768 MW;  4CDD1913 CRC32;

Query Match         14.8%;  Score 86;  DB 5;  Length 1059;
Best Local Similarity   27.2%;  Pred. No. 0.29%;
Matches   31;  Conservative  12;  Mismatches  33;  Indels  38;  Gaps   9;

Qy    20 TLCSMEEAINE------RIQEVAGSLIFRAISSIGLECQSVTSRGDLATCPRGFA-VTG 71
         |:|  ||  ||          |  |::  | ::|  |||   ||  | :|:
```

```
Db    80 TICGGEEIVSRGQCEVVCRSRLTGANFLWRV--DDGTPCQAATSR--AVCSKGSCQIVG 134
Qy    72 C-------TCGSACGS----------WDVRAE-TTCHCQCAG-MDWTGA 101
           ||  ||                      ||  ||||  |||  :|:||
Db   135 CDGLISSSFRFDACG-VCGGRGDTCDNGKFIWKVSEEYTACASNCDDIVDWSGA 187

RESULT 2
Q90285
ID   Q90285           PRELIMINARY;       PRT;   1476 AA.
AC   Q90285; Q98847;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   NOTCH-3 HOMOLOG (FRAGMENT).
OS   Carassius auratus (Goldfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC   Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC   Cyprinoidea; Cyprinidae; Cyprininae; Carassius.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=RETINA (10D POST-OPTIC NERVE CRUSH);
RX   MEDLINE: 97359070.
RA   SULLIVAN S.A., BARTHEL L.K., LARGENT B.L., RAYMOND P.A.;
RT   "A goldfish Notch-3 homologue is expressed in neurogenic regions of
RT   embryonic, adult, and regenerating brain and retina.";
RL   Dev. Genet. 20:208-223(1997).
DR   EMBL; U09191; AAB17010.2; -.
DR   HSSP; P01132; 1EPH.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00008; EGF; 11.
DR   PFAM; PF00066; notch; 3.
FT   NON_TER    1      1
FT   NON_TER 1476   1476
SQ   SEQUENCE 1476 AA; 160385 MW; D6077129 CRC32;

Query Match           14.5%;  Score 84.5;  DB 13;  Length 1476;
Best Local Similarity 33.3%;  Pred. No. 0.61;
Matches   28;  Conservative   7;  Mismatches  32;  Indels 17;  Gaps 5;

Qy   36 AGSLIFRAISSIGLECQ-----SVTSRGDLATCPRGFAVTGCT------CGS-ACGSWDV 83
           ||: |:::|  ||||                :||:|| ||  ||      :|
Db  357 AGANCERSMNCKELPCYNGGSCSDTSRGARCTCIQGFGGPQQCHRSNEGCSSKPCSNGGL 416

Qy   84 RAETTC----HCQCAGMDWTGARC 103
           |  |||    ||||
Db  417 CTEETSPFFHCQCIS-GWKGIRC 439

RESULT 3
O42373
ID   O42373           PRELIMINARY;       PRT;   762 AA.
AC   O42373;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   NOTCH RECEPTOR PROTEIN (FRAGMENT).
GN   NOTCH5.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC   Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC   Cyprinoidea; Cyprinidae; Rasborinae; Danio.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   WESTIN J., LARDELLI M.;
RL   Dev. Genes Evol. 207:51-63(1997).
DR   EMBL; Y10353; CAA71379.1; -.
DR   HSSP; P00740; 1IXA.
DR   ZFIN; ZDB-GENE-980526-78; notch5.
DR   PFAM; PF00023; ank; 3.
DR   PFAM; PF00008; EGF; 5,
DR   PFAM; PF00066; notch; 3.
```

```
KW   Glycoprotein.
FT   NON_TER    1      1
FT   NON_TER  762    762
SQ   SEQUENCE 762 AA; 83560 MW; 9FEIF01E CRC32;

Query Match           14.2%;  Score 82.5;  DB 13;  Length 762;
Best Local Similarity 32.1%;  Pred. No. 0.5;
Matches   27;  Conservative   9;  Mismatches  31;  Indels 17;  Gaps 5;

Qy   36 AGSLIFRAISSIGLECQ-----SVTSRGDLATCPRGFAVTGCT------CGS-ACGSWDV 83
           ||| |:::|                :|:||     ||  ||      ||| :
Db  142 AGSNCERSMNCKELPCYNGGSCTLTTRGARCTCIQGFGGPLCQHRSNDGCSSKPCHNGGL 201

Qy   84 RAETTC----HCQCAGMDWTGARC 103
           |  |||    |||| ||
Db  202 CTEETSYPFFHCQCTN-GWKGKRC 224

RESULT 4
O88281
ID   O88281           PRELIMINARY;       PRT;   1574 AA.
AC   O88281;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MEGF6.
GN   MEGF6.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE: 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple EGF-
RT   like motifs by motif-trap screening.";
RL   Genomics 51:27-34(1998).
DR   EMBL; AB011532; BAA32462.1; -.
DR   HSSP; P00736; 1APQ.
DR   PROSITE; PS00010; ASX_HYDROXYL; 5.
DR   PROSITE; PS01187; EGF_CA; 5.
DR   PFAM; PF00008; EGF; 25.
KW   Glycoprotein; EGF-like domain.
SQ   SEQUENCE 1574 AA; 165445 MW; C49E6EA6 CRC32;

Query Match           14.2%;  Score 82.5;  DB 11;  Length 1574;
Best Local Similarity 35.6%;  Pred. No. 1.1;
Matches   21;  Conservative   3;  Mismatches   26;  Indels  9;  Gaps 2;

Qy   47 IGLECQSVTSRGDLATCPRGFAVTGCTCGSACGSWDVRAETTCHCQCAGMDWTGARCCR 105
           ||  || ||:|||||  || :|| ||: |  ||| |   |:||  |||  ||| :|
Db  813 VGSRCQD-------TCSAGWYGTGCQIRCACANDGHCDPTTGRCSCA-PGWTGLSCQR 862

RESULT 5
Q9Y6L8
ID   Q9Y6L8           PRELIMINARY;       PRT;   2321 AA.
AC   Q9Y6L8;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   NOTCH3.
GN   NOTCH3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   GUNEL M., ARTAVANIS-TSAKONAS S.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
```

Db   182 NKSCDHITGEILCRT-GYIGLTCEH--------PCPAGLYGPGCKLKCNCEHGGECNHVT 232

Qy   89 CHCQCAGMDWTGARC 103
          ||| |||| |
Db   233 GQCQCL-PGWTGSNC 246

RESULT   7
Q17494
ID Q17494   PRELIMINARY;   PRT; 1106 AA.
AC Q17494;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE B0393.5 PROTEIN.
GN B0393.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RL Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z37983; CAA86058.1; -.
DR HSSP; P07204; 2ADX.
DR PROSITE; PS00010; ASX_HYDROXYL; 6.
DR PROSITE; PS01187; EGF_CA; 5.
DR PFAM; PF00008; EGF; 5.
KW Glycoprotein; EGF-like domain.
SQ SEQUENCE 1106 AA; 121137 MW; 2C93CB51 CRC32;

Query Match   13.4%; Score 78; DB 5; Length 1106;
Best Local Similarity   31.1%; Pred. No. 2.3;
Matches 19; Conservative   4; Mismatches 22; Indels 16; Gaps   3;

Qy   59 DLATCPRG-FAVTGCTCGSACGSWDVRAETTCHC--------QCAG---MDWTGAR 102
          || ||| | | : ||| :: |||| | |:   :||||
Db   772 DLQTCPPGRLGEPLCDRECAAGHYGINCESTCHCDGSVACDVITGMCPGALCRAGWEGSS 831

Qy   103 C 103
Db   832 C 832

RESULT   8
O88671
ID O88671   PRELIMINARY;   PRT; 589 AA.
AC O88671;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE DELTA 3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]

DR EMBL; AF058900; AAC14346.1; -.
DR EMBL; AF058881; AAC14346.1; JOINED.
DR EMBL; AF058882; AAC14346.1; JOINED.
DR EMBL; AF058883; AAC14346.1; JOINED.
DR EMBL; AF058884; AAC14346.1; JOINED.
DR EMBL; AF058885; AAC14346.1; JOINED.
DR EMBL; AF058886; AAC14346.1; JOINED.
DR EMBL; AF058887; AAC14346.1; JOINED.
DR EMBL; AF058888; AAC14346.1; JOINED.
DR EMBL; AF058889; AAC14346.1; JOINED.
DR EMBL; AF058890; AAC14346.1; JOINED.
DR EMBL; AF058891; AAC14346.1; JOINED.
DR EMBL; AF058892; AAC14346.1; JOINED.
DR EMBL; AF058893; AAC14346.1; JOINED.
DR EMBL; AF058894; AAC14346.1; JOINED.
DR EMBL; AF058895; AAC14346.1; JOINED.
DR EMBL; AF058896; AAC14346.1; JOINED.
DR EMBL; AF058897; AAC14346.1; JOINED.
DR EMBL; AF058898; AAC14346.1; JOINED.
DR EMBL; AF058899; AAC14346.1; JOINED.
DR PROSITE; PS01187; EGF_CA; 16.
DR PROSITE; PS00010; ASX_HYDROXYL; 18.
KW Glycoprotein; EGF-like domain.
SQ SEQUENCE 2321 AA; 243628 MW; 98AA7CB8 CRC32;

Query Match   14.1%; Score 82; DB 4; Length 2321;
Best Local Similarity   29.9%; Pred. No. 1.8;
Matches 23; Conservative   7; Mismatches 23; Indels 24; Gaps   4;

Qy   47 IGLECQSVTSRGDLATCPRGFAVTGCTCGSACGSWDVRAETTC---------- 90
          :|: ||| | :||||: :| ||||| : | | :: |||||
Db 1300 VGVPCQQ-TPRGPRCACPPG--LSGPSCRSFPGSPPGASNASCAAAPCLHGGSCRPAPLA 1356

Qy   90 ---HCQCAGMDWTGARC 103
             | | ||||||
Db 1357 PFFRCACA-QGWTGPRC 1372

RESULT   6
Q9Y151
ID Q9Y151   PRELIMINARY;   PRT; 594 AA.
AC Q9Y151;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE BCDNA.GH03529.
GN BCDNA.GH03529.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RA RUBIN G.M., WAN K.H., HARVEY D., LEWIS S.E., BROKSTEIN P., TSANG G.,
RA AGBAYANI A., ARCAINA T.T., BAXTER E., BLAZEJ R.G., BUTENHOFF C.,
RA CHAMPE M., CHAVEZ C., CHEW M., DOYLE C.M., FARFAN D.E., FRISE E.,
RA GALLE R., GEORGE R.A., HARRIS N.L., HOSKINS R.A., EVANS-HOLM M.,
RA HOUSTON K.A., HUMMASTI S.R., KIM E., LI P., MOSHREFI M., PACLEB J.M.,
RA PARK S., SEQUEIRA A., SETHI H., SNIR E., SVIRSKAS R.R., WEINBURG T.,
RA CELNIKER S.E.;
RT "Full length Drosophila melanogaster cDNA sequence.";
RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF145617; AAD38592.1; -.
SQ SEQUENCE 594 AA; 66234 MW; CD1F8716 CRC32;

Query Match   13.4%; Score 78; DB 5; Length 594;
Best Local Similarity   28.0%; Pred. No. 1.2;
Matches 21; Conservative   6; Mismatches 38; Indels 10; Gaps   3;

Qy   29 NERIQEVAGSLIFRAISSIGLECQSVTSRGDLATCPRGFAVTGCTCGSACGSWDVRAETT 88
          |:   : : | | ||| || || ||

```
RP   SEQUENCE FROM N.A.
RA   BOULTER J., GREENFIELD A., WEINMASTER G.;
RT   "Rattus norvegicus mRNA for Delta 3: a putative ligand for Notch.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF084576; AAC33303.1; -.
DR   HSSP; P00740; 1IXA.
DR   PFAM; PF00008; EGF; 5.
KW   Glycoprotein.
SQ   SEQUENCE   589 AA;  61424 MW;  810A3749 CRC32;

Query Match             13.2%;  Score 76.5;  DB 11;  Length 589;
Best Local Similarity   28.8%;  Pred. No. 1.7;
Matches   34;  Conservative    9;  Mismatches   42;  Indels   33;  Gaps  8;

Qy   21 LC----SMEEAINERIQEVAGS--LIFRAISSIGLECQ-----SVTSRGDLATCPRGF-- 68
         ||    |    :    :  || :  : || :       | :      ||||||||||
Db  249 LCTVPVSTSSCLNSRVSGPAGTGCLLPGPGPCDGNPCANGGSCSETPGSFECACPRGFYG 308

Qy   68 ---AVTGCTC------GSAC-GSWDVRAETTCHCQCAGMDWTGARC------CRVQP 108
         :|:||   ||     ||| |   | : : :|| | :     :||
Db  309 PRCEVSGVTCADGPCFNGGLCVGGEDPDSAVCHCPPA---FQGSNCERRVDRCSLQP 363

RESULT 9
ID   O13149
AC   O13149;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   NOTCH 2 (FRAGMENT).
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC   Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
OC   Tetraodontiformes; Tetraodontoidei; Tetraodontidae; Fugu.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   NAKAMURA T., TROWSDALE J.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB004829; BAA20535.1; -.
DR   HSSP; P00740; 1IXA.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS01187; EGF_CA; 20.
DR   PFAM; PF00023; ank; 6.
DR   PFAM; PF00008; EGF; 35.
DR   PFAM; PF00066; notch; 3.
KW   Glycoprotein; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   2447 AA;  262542 MW;  3CDA4F7A CRC32;

Query Match             13.1%;  Score 76;  DB 13;  Length 2447;
Best Local Similarity   26.7%;  Pred. No. 8.7;
Matches   24;  Conservative    2;  Mismatches   20;  Indels   44;  Gaps  6;

Qy   50 ECQSVTSRGDLATCPRGFAVTGCTC---------------GSACG-----SW 81
         || :   |   ||| :       |||               :||
Db  982 ECDSQPCKNG-GTCTDGLGTYRCTCPAGYNGQNCQNYVNLCRQVRCHNGGSCSHTGATSW 1040

Qy   82 DVRAETTCHCQCAGMDWTGARC-----CR 105
         ||||        |        ||     ||
Db 1041 ------TCHCt--MGWTGPYCDVPDMSCR 1061

RESULT 10
ID   O75095
AC   O75095;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MEGF6 (FRAGMENT).
```

```
GN   MEGF6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple EGF-
RT   like motifs by motif-trap screening.";
RL   Genomics 51:27-34(1998).
DR   EMBL; AB011539; BAA32467.1; -.
DR   HSSP; P02468; 1KLO.
DR   PFAM; PF00008; EGF; 2.
FT   NON_TER       1      1
SQ   SEQUENCE   153 AA;  15579 MW;  1AC64E20 CRC32;

Query Match             13.0%;  Score 75.5;  DB 4;  Length 153;
Best Local Similarity   35.7%;  Pred. No. 0.54;
Matches   20;  Conservative    1;  Mismatches   26;  Indels    9;  Gaps  3;

Qy   48 GLECQSVTSRGDLATCPRGFAVTGCTCGSACGSWDVRAETTCHCQCAGMDWTGARC 103
         | :|  |||  |||| |      :|| ||         :||  || |    |  | 
Db   51 GAPCDPVTG---LCLCPPG--RSGATCNLDCRRGQFGPSCTLHCDCGG----GADC 97

RESULT 11
ID   Q06007
AC   Q06007;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   NOTCH PROTEIN HOMOLOG 1 (MOTCH A PROTEIN) (FRAGMENT).
GN   NOTCH1 OR MOTCH A.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F1 (CBA X C57BL); TISSUE=WHOLE EMBRYO;
RX   MEDLINE; 93178563.
RA   LARDELLI M., LENDAHL U.;
RT   "Motch A and motch B--two mouse Notch homologues coexpressed in a wide
RT   variety of tissues.";
RL   Exp. Cell Res. 204:364-372(1993).
DR   EMBL; X69278; CAA48339.1; -.
DR   HSSP; P00743; 1APO.
DR   MGD; MGI:97363; Notch1.
DR   PFAM; PF00008; EGF; 6.
KW   Differentiation; Neurogenesis; Repeat.
FT   NON_TER       1      1
FT   NON_TER     387    387
SQ   SEQUENCE   387 AA;  41497 MW;  D1FD6C00 CRC32;

Query Match             13.0%;  Score 75.5;  DB 11;  Length 387;
Best Local Similarity   29.1%;  Pred. No. 1.4;
Matches   25;  Conservative    4;  Mismatches   24;  Indels   33;  Gaps  5;

Qy   48 GLECQSV-----------------------TSRGDLATCPRGFAVTGCTC---GSACGSWDVR 84
         | :| ||                       ||| :  : |  | :||    |  | 
Db  141 GRRCESVINGCRGKPCKNGGVCAVASNTARGFICRCPAGF--EGATCENDARTCGSLRCL 198

Qy   85 AETTC------HCQCAGMDWTGARC 103
         | :||      |       |   || 
Db  199 NGGTCISGPRSPTCLCLG-SFTGPEC 223

RESULT 12
```

Db   305 LRCEVSGVTCADGPCFNGGLCVGGEDPDSXYVCHCP---PGFQGSNCEKRVDRCSLQP 359

RESULT 14
O88516
ID   O88516   PRELIMINARY;   PRT;   592 AA.
AC   O88516;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   DELTA-LIKE 3 ALTERNATE SPLICE FORM 2.
GN   DLL3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SVJ;
RX   MEDLINE; 98324780.
RA   KUSUMI K., SUN E.S., KERREBROCK A.W., BRONSON R.T., CHI D.C.,
RA   BULOTSKY M.S., SPENCER J.B., BIRREN B.W., FRANKEL W.N., LANDER E.S.;
RT   "The mouse pudgy mutation disrupts Delta homologue Dll3 and initiation
RT   of early somite boundaries.";
RL   Nat. Genet. 19:274-278(1998).
DR   EMBL; AF068865; AAC40170.1; -.
DR   HSSP; P00740; 1IXA.
DR   PFAM; PF00008; EGF; 6.
DR   PRINTS; PR00010; EGFBLOOD.
KW   Glycoprotein.
SQ   SEQUENCE   592 AA;   62069 MW;   C1829CC9 CRC32;

Query Match   12.8%;
Best Local Similarity   28.8%;   Pred. No. 2.9;   DB 11;   Length 592;
Matches   34; Conservative   8; Mismatches   43; Indels   33; Gaps   8;

Qy   21 LC----SMEEAINERIQEVA--GSLIFRAISSIGLECQ-----SVTSRGDLATCPRGF-- 68
                                                              ||||
Db   245 LCTVPVSTSSCLNSRVPGPASTGCLLPGPGPCDGNPCANGGSCSETSGSFECACPRGFYG 304

Qy   68 --AVTGCTC------GSAC-GSWDVRAETTCHCQCAGMDWTGARC-----CRVQP 108
            || | ||| | ||| | ||
Db   305 LRCEVSGVTCADGPCFNGGLCVGGEDPDSAYVCHCP---PGFQGSNCEKRVDRCSLQP 359

RESULT 15
Q94438
ID   Q94438   PRELIMINARY;   PRT; 1698 AA.
AC   Q94438;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   185 KDA SILK PROTEIN.
GN   SP185.
OS   Chironomus pallidivittatus (Midge).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC   Chironomoidea; Chironomidae; Chironominae; Chironomus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SALIVARY GLAND;
RA   CASE S.T., COX C., BELL W.C., HOFFMAN R.T., MARTIN J., HAMILTON R.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U54640; AAA99803.1; -.
DR   HSSP; P18055; 2MRB.
DR   PRINTS; PR00876; MTNEMATODE.
SQ   SEQUENCE 1698 AA; 186164 MW; 85E8E520 CRC32;

Query Match   12.8%;
Best Local Similarity   26.4%;   Pred. No. 8.7;   DB 5;   Length 1698;
Matches   23; Conservative   6; Mismatches   25; Indels   33; Gaps   5;

O96388
ID   O96388   PRELIMINARY;   PRT;   75 AA.
AC   O96388;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   METALLOTHIONEIN.
OS   Perna viridis.
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytilidae; Perna.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KHOO H.W., WONG Y.W., PATEL K.H.;
RT   "Green mussel (Perna viridis) metallothionein cDNA.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF036904; AAD02054.1; -.
SQ   SEQUENCE   75 AA;   7415 MW;   942586D8 CRC32;

Query Match   12.8%;
Best Local Similarity   27.9%;   Pred. No. 0.33;   DB 5;   Length 75;
Matches   17; Conservative   9; Mismatches   18; Indels   17; Gaps   4;

Qy   51 CQSVTSRGDLATCPRGFAVTGCTCGSACGSWDVRAETTCHCQCAGMDWTGARC---CRVQ 107
          | :| : |:: || || || | |:| | |||:| |:|
Db   5 CNCIETQ-VCICGTGCSGEGCRCGDAC-----KCSSGGCGCSG------CKVVCKCQ 50

Qy   108 P 108
         |
Db   51 P 51

RESULT 13
O35675
ID   O35675   PRELIMINARY;   PRT;   585 AA.
AC   O35675;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   M-DELTA-LIKE 3 GENE PRECURSOR.
GN   DLL3 OR M-DELTA-LIKE 3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57B16 X DBA; TISSUE=PRIMITIVE STREAK;
RX   MEDLINE; 97417575.
RA   DUNWOODIE S.L., HENRIQUE D.M.P., HARRISON S.M., BEDDINGTON R.S.P.;
RT   "Mouse Dll3: a novel divergent Delta gene which may complement the
RT   function of other Delta homologues during early pattern formation in
RT   the mouse embryo.";
RL   Development 124:3065-3076(1997).
DR   EMBL; Y11895; CAA72637.1; -.
DR   HSSP; P00740; 1IXA.
DR   MGD; MGI:1096877; Dll3.
DR   PFAM; PF00008; EGF; 6.
DR   PRINTS; PR00010; EGFBLOOD.
KW   Signal; Glycoprotein.
FT   SIGNAL   1   32   POTENTIAL.
SQ   SEQUENCE   585 AA;   61129 MW;   0F5E6187 CRC32;

Query Match   12.8%;
Best Local Similarity   28.8%;   Pred. No. 2.8;   DB 11;   Length 585;
Matches   34; Conservative   8; Mismatches   43; Indels   33; Gaps   8;

Qy   21 LC----SMEEAINERIQEVA--GSLIFRAISSIGLECQ-----SVTSRGDLATCPRGF-- 68
          || || |:: || | : | |||| |||||
Db   245 LCTVPVSTSSCLNSRVPGPASTGCLLPGPGPCDGNPCANGGSCSETSGSFECACPRGFYG 304

Qy   68 --AVTGCTC------GSAC-GSWDVRAETTCHCQCAGMDWTGARC-----CRVQP 108
            || | ::| | ::| | | :||

```
Qy     52  QSVTSRGDLATCPRGFAVTGCT--CGSAC----------GSWDVRAETTCHCQC--- 94
           |:|  |:  :||   ||         :||  |::|·|||
Db   1355  QTWNSQTCQCSCPASGTCTGAQVWCSKACKCVCPAQKKCDSPKTWD---ESSCSCQCPKN 1411

Qy     94  --------AGMDWTGARC---CRVQP 108
           ||--|--|||·|-|--
Db   1412  MRPPKGGCNAGRTWDDATCSEKCAAVP 1438


Search completed: May 26, 2000, 03:19:34
Job time: 10004 sec
```

OM protein - protein search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on:　May 26, 2000, 00:06:17 ; Search time 39.3 Seconds
(without alignments)
65.092 Million cell updates/sec

Title:　　　　US-09-099-898-2
Perfect score:　581
Sequence:　　1 MKALCLLLPVLGLLVSSKT.........CHCQCAGMDWTGARCCRVQP 108

Scoring table:　BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:　188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters:　188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :　A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	581	100.0	108	1	W87710	A cysteine rich so
2	581	100.0	109	1	Y12933	Amino acid sequenc
3	318	54.7	114	1	W87707	A cysteine rich so
4	315	54.2	114	1	W87708	A cysteine rich so
5	284.5	49.0	111	1	W87706	A cysteine rich so
6	273.5	47.1	105	1	W87705	A cysteine rich so
7	245.5	42.3	111	1	Y12613	A cysteine rich so
8	227	39.1	52	1	W87704	Human 5' EST secre
9	202.5	34.9	111	1	W68510	A cysteine rich so
10	82	14.1	1872	1	W49698	Partial human Notc
11	82	14.1	2321	1	R98207	Human Notch3 prote
12	76.5	13.2	60	1	W74444	Nucleotide used in
13	73	12.6	131	1	R91428	Mouse Tie receptor
14	72	12.4	1111	1	R91427	Kalinin/laminin 5
15	72	12.4	1193	1	W30826	Kalinin/laminin 5
16	71.5	12.3	519	1	W67738	The novel tyrosina
17	71.5	12.3	2476	1	R27199	Pig p105 zona pell
18	71	12.2	397	1	R39821	ALP of Lysobacter
19	69	11.9	1094	1	Y06816	Truncated tie rece
20	68.5	11.8	2471	1	W24566	Human Notch2 (humN
21	68	11.7	102	1	R73954	Serine protease C-
22	68	11.7	1122	1	R39820	Human tie tyrosine
23	68	11.7	1138	1	Y13350	tie receptor kinas
24	67	11.5	915	1	W50895	Amino acid sequenc
25	67	11.5	1801	1	R67757	Rat laminin B2 cha
26	66.5	11.4	488	1	R67758	Human fetal brain
27	66.5	11.4	524	1	R67759	Human fetal brain
28	66.5	11.4	670	1	R75352	Human fetal brain
29	66.5	11.4	769	1	R40167	Human fetal brain
30	66	11.4	177	1	W20066	Recombinant growth
31	66	11.4	297	1	P50296	Thielavia terrestr
32	65.5	11.3	1080	1	W95557	Mouse epidermal gr
33	65.5	11.3	1964	1	R52633	Mus musculus notch
34	65	11.2	289	1		Guinea pig PH-30,
35	65	11.2	1148	1	W87895	Human JAGGED2 prot
36	65	11.2	2707	1	W27161	Mouse receptor ME2
37	64.5	11.1	375	1	W44141	Thermostable alkal
38	64.5	11.1	2799	1	W81867	Human tumour suppr
39	64	11.0	197	1	W36951	Protein encoded by
40	64	11.0	213	1	R45359	Wheat germ aggluti
41	64	11.0	500	1	W94496	Human delta-2 prot
42	64	11.0	659	1	W94497	Human delta-2 matu
43	64	11.0	685	1	W80813	Nucleotide sequenc
44	64	11.0	685	1	W94507	Human delta-2 prot
45	64	11.0	688	1	P80361	Tick WGL+ antigen.

ALIGNMENTS

RESULT　1
W87710
ID　W87710 standard; Protein; 108 AA.
AC　W87710;
DT　09-MAR-1999　(first entry)
DE　A cysteine rich soluble protein designated C23.
KW　Cysteine rich soluble protein; CRSP; C23; cell development;
KW　mammalian immune system; antibody; abnormal proliferation; cancer;
KW　inflammation; degeneration; regeneration; degeneration: atrophy.
OS　Homo sapiens.
FH　Key　　　　　Location/Qualifiers
FT　Protein　　　1..18
FT　　　　　　　/note= "mature protein"
PN　WO9858061-A1.
PD　23-DEC-1998.
PF　18-JUN-1998; U12236.
PR　09-OCT-1997; US-061641.
PR　19-JUN-1997; US-878730.
PR　19-JUN-1997; US-878878.
PA　(SCHE) SCHERING CORP.
PI　Franz-Bacon K, Gorman DM, McClanahan TK;
DR　WPI; 99-095339/08.
DR　N-PSDB; V84059.
PT　New cysteine-rich soluble proteins - used to modulate proliferation,
PT　differentiation, trafficking and development of cells, e.g. for
PT　treating inflammation, cancer and degeneration
PS　Claim 1; Page 18-19; 119pp; English.
CC　The present sequence represents a cysteine rich soluble protein (CRSP)
CC　designated C23. CRSP proteins, and their (ant)agonists, are used to
CC　modulate physiology, differentiation, trafficking and development of
CC　of cells (including those in culture), particularly cells of the
CC　mammalian immune system. They are used for treatment of abnormal
CC　proliferation (cancer, inflammation or degeneration), regeneration,
CC　degeneration and atrophy. The proteins are also used to raise, or
CC　detect, antibodies, to design oligonucleotides for library screening,
CC　in drug screens and to isolate cognate receptors. The antibodies are
CC　used for affinity purification of CRSP, to screen expression libraries,
CC　to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC　to produce anti-idiotypic antibodies (useful for diagnosis), in
CC　competitive drug screens, and as therapeutic modulators.
SQ　Sequence　108 AA;

Query Match　　　　　　　100.0%;　Score 581; DB 1; Length 108;
Best Local Similarity　100.0%;　Pred. No. 8.3e-50;
Matches 108;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　1 MKALCLLLPVLGLLVSSKTLCSMEEAINERIQEVAGSLIFRAISSIGLECQSVTSRGDL 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　1 MKALCLLLPVLGLLVSSKTLCSMEEAINERIQEVAGSLIFRAISSIGLECQSVTSRGDL 60

Qy 61 ATCPRGFAVTGCTCGSACGSWDVRAETTCHCQCAGMDWTGARCCRVQP 108
       ||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 ATCPRGFAVTGCTCGSACGSWDVRAETTCHCQCAGMDWTGARCCRVQP 108

RESULT　2

```
Y12933
ID   Y12933 standard; Protein; 109 AA.
AC   Y12933;
DT   17-JUN-1999  (first entry)
DE   Amino acid sequence of a human secreted peptide.
KW   Human secreted protein; cancer; immune disorder; infection;
KW   inflammatory disorder; skin disorder; tumour; atherosclerosis;
KW   restenosis; autoimmune disorder; Alzheimer's disease;
KW   peripheral neuropathy; trauma; spinal cord injury; allergy;
KW   hematopoietic disorder; skeletal disorder; neurological disorder;
KW   arthritic disorder; asthma; immunodeficiency disease; AIDS;
KW   transplant rejection; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /note= "signal peptide"
FT   Protein         24..108
FT                   /note= "secreted protein"
PN   WO9911293-A1.
PD   11-MAR-1999.
PF   03-SEP-1998; U18360.
PR   12-SEP-1997; US-058974.
PR   05-SEP-1997; US-057626.
PR   05-SEP-1997; US-057663.
PR   05-SEP-1997; US-057669.
PR   12-SEP-1997; US-058656.
PR   12-SEP-1997; US-058667.
PR   12-SEP-1997; US-058973.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Brewer LA, Ebner R, Lafleur DW, Moore PA, Olsen HS,
PI   Rosen GA, Ruben SM, Shi Y;
DR   WPI; 99-204988/17.
DR   N-PSDB; X51720.
PT   New isolated human genes and the secreted polypeptides they encode
PT   - useful for diagnosis and treatment of e.g. neurological disorders,
PT   tumours, immune disorders, inflammation or haematological disorders
PS   Claim 11; Page 190; 215pp; English.
CC   Y12914-68 represent human secreted proteins. The polypeptides and
CC   their corresponding polynucleotides are useful for preventing,
CC   treating or ameliorating medical conditions, e.g. by protein or gene
CC   therapy. Pathological conditions can also be diagnosed by determining
CC   the amount of the new polypeptides in a sample or by determining the
CC   presence of mutations in the new polynucleotides. Specific uses are
CC   described for each polynucleotide, based on which tissues they are
CC   most highly expressed in, and include developing products for the
CC   diagnosis or treatment of cancer, immune disorders, infection,
CC   inflammatory disorders, skin disorders, tumours, atherosclerosis,
CC   restenosis, autoimmune disorders, Alzheimer's disease, peripheral
CC   neuropathies, trauma, spinal cord injuries, allergy, hematopoietic
CC   disorders, skeletal disorders, neurological disorders, arthritic
CC   disorders, asthma, immunodeficiency diseases, AIDS and transplant
CC   rejection. The polypeptides are also useful for identifying their
CC   binding partners.
SQ   Sequence 109 AA;

Query Match            100.0%;  Score 581;  DB 1;  Length 109;
Best Local Similarity  100.0%;  Pred. No. 8.4e-50;
Matches  108;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1 MKALCLLLPVLGLLVSSKTLCSMEEAINERIQEVAGSLIFRAISSIGLECQSVTSRGDL 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MKALCLLLPVLGLLVSSKTLCSMEEAINERIQEVAGSLIFRAISSIGLECQSVTSRGDL 60

Qy    61 ATCPRGFAVTGCTCGSACGSWDVRAETTCHCQCAGMDWTGARCCRVQP 108
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ATCPRGFAVTGCTCGSACGSWDVRAETTCHCQCAGMDWTGARCCRVQP 108

RESULT  3
W87707
ID   W87707 standard; Protein; 114 AA.
AC   W87707;
DT   09-MAR-1999  (first entry)
DE   A cysteine rich soluble protein designated C19.
KW   Cysteine rich soluble protein; CRSP; C19; cell development;
KW   mammalian immune system; antibody; abnormal proliferation; cancer;
KW   inflammation; degeneration; regeneration; atrophy.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   Protein         1..20
FT                   /note= "mature protein"
PN   WO9858061-A1.
PD   23-DEC-1998.
PF   18-JUN-1998; U12236.
PR   09-OCT-1997; US-061641.
PR   19-JUN-1997; US-878730.
PR   19-JUN-1997; US-878878.
PA   (SCHE ) SCHERING CORP.
PI   Franz-Bacon K, Gorman DM, McClanahan TK;
DR   WPI; 99-095339/08.
DR   N-PSDB; V84056.
PT   New cysteine-rich soluble proteins - used to modulate proliferation,
PT   differentiation, trafficking and development of cells, e.g. for
PT   treating inflammation, cancer and degeneration
PS   Claim 1; Page 15; 119pp; English.
CC   The present sequence represents a cysteine rich soluble protein (CRSP)
CC   designated C19. CRSP proteins, and their (ant)agonists, are used to
CC   modulate physiology, differentiation, trafficking and development
CC   of cells (including those in culture), particularly cells of the
CC   mammalian immune system. They are used for treatment of abnormal
CC   proliferation (cancer, inflammation or degeneration), regeneration,
CC   degeneration and atrophy. The proteins are also used to raise, or
CC   detect, antibodies, to design oligonucleotides for library screening,
CC   in drug screens and to isolate cognate receptors. The antibodies are
CC   used for affinity purification of CRSP, to screen expression libraries,
CC   to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC   to produce anti-idiotypic antibodies (useful for diagnosis), in
CC   competitive drug screens, and as therapeutic modulators.
SQ   Sequence 114 AA;

Query Match            54.7%;  Score 318;  DB 1;  Length 114;
Best Local Similarity  55.8%;  Pred. No. 3.4e-24;
Matches  58;  Conservative  16;  Mismatches  30;  Indels  0;  Gaps  0;

Qy     4 LCLLLPVLGLLVSSKTLCSMEEAINERIQEVAGSLIFRAISSIGLECQSVTSRGDLATC 63
          | |||| | ||||||||: :|:|| ||::::  ||  ||:  |  |||| ::|:| ||||
Db     8 LLFLFFLVPELLGSSMPLCPIDEAIDKKKQDFNSLFPNAIKNIGLNCWTVSSRGKLASC 67

Qy    64 PRGFAVTGCTCGSACGSWDVRAETTCHCQCAGMDWTGARCCRVQ 107
          | || :||||||||||| || :|| :|||| :||| ||||::|
Db    68 PEGTAVLSCSCGSACGSWDIREEKVCHCQCCARIDWTAARCCKLQ 111

RESULT  4
W87708
ID   W87708 standard; Protein; 114 AA.
AC   W87708;
DT   09-MAR-1999  (first entry)
DE   A cysteine rich soluble protein designated C19.
KW   Cysteine rich soluble protein; CRSP; C19; cell development;
KW   mammalian immune system; antibody; abnormal proliferation; cancer;
KW   inflammation; degeneration; regeneration; atrophy.
OS   Rattus sp.
FH   Key             Location/Qualifiers
FT   Protein         1..17
FT                   /note= "mature protein"
PN   WO9858061-A1.
PD   23-DEC-1998.
PF   18-JUN-1998; U12236.
PR   09-OCT-1997; US-061641.
PR   19-JUN-1997; US-878730.
PR   19-JUN-1997; US-878878.
PA   (SCHE ) SCHERING CORP.
PI   Franz-Bacon K, Gorman DM, McClanahan TK;
```

```
DR   WPI; 99-095339/08.
DR   N-PSDB; V84057.
PT   New cysteine-rich soluble proteins - used to modulate proliferation,
PT   differentiation, trafficking and development of cells, e.g. for
PT   treating inflammation, cancer and degeneration
PS   Claim 16; Page 16; 119pp; English.
CC   The present sequence represents a cysteine rich soluble protein (CRSP)
CC   designated C19. CRSP proteins, and their (ant)agonists, are used to
CC   modulate physiology, differentiation, trafficking and development
CC   of cells (including those in culture), particularly cells of the
CC   mammalian immune system. They are used for treatment of abnormal
CC   proliferation (cancer, inflammation or degeneration), regeneration,
CC   degeneration and atrophy. The proteins are also used to raise, or
CC   detect, antibodies, to design oligonucleotides for library screening,
CC   in drug screens and to isolate cognate receptors. The antibodies are
CC   used for affinity purification of CRSP, to screen expression libraries,
CC   to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC   to produce anti-idiotypic antibodies (useful for diagnosis), in
CC   competitive drug screens, and as therapeutic modulators.
SQ   Sequence   114 AA;

Query Match        54.2%;     Score 315; DB 1; Length 114;
Best Local Similarity  54.1%;  Pred. No. 6.6e-24;
Matches   60; Conservative   16; Mismatches  31; Indels   4; Gaps   1;

Qy    1 MKALCLLL---PVLGLLVSSKTLCSMEEAINERIQEVAGSLIFRAISSIGLECQSVTS 56
         | ||  ||| |||| | | |:|||||||||||||||:||: |: |||:| | ||:|
Db    1 MKNLSFLLFLFIVLGLLGPSMSLCPMDEAISKKINQDFSSLLPAAMKNTVLHCWSVSS 60

Qy   57 RGDLATCPRGFAVTGCTCGSACGSWDVRAETTCHCQCAGMDWTGARCCRVQ 107
         ||:||:||  ||||| | |||| |||||||||||:|||||||||||| ::
Db   61 RGRLASCPEGTTVTSCSCGSGCGSWDVREDTMCHCQCGSIDWTAARCCTLR 111

RESULT   5
W87709
ID   W87709 standard; Protein; 111 AA.
AC   W87709;
DT   09-MAR-1999 (first entry)
DE   A cysteine rich soluble protein designated C10.
KW   Cysteine rich soluble protein; CRSP; C10; cell development;
KW   mammalian immune system; antibody; abnormal proliferation; cancer;
KW   inflammation; degeneration; regeneration; degeneration; atrophy.
OS   Homo sapiens.
FH   Key         Location/Qualifiers
FT   Protein     1..20
FT               /note= "mature protein"
PN   WO9858061-A1.
PD   23-DEC-1998.
PF   18-JUN-1998; U12236.
PR   09-OCT-1997; US-061641.
PR   19-JUN-1997; US-878730.
PR   19-JUN-1997; US-878878.
PA   (SCHE ) SCHERING CORP.
PI   Franz-Bacon K, Gorman DM, McClanahan TK;
DR   WPI; 99-095339/08.
DR   N-PSDB; V84058.
PT   New cysteine-rich soluble proteins - used to modulate proliferation,
PT   differentiation, trafficking and development of cells, e.g. for
PT   treating inflammation, cancer and degeneration
PS   Claim 1; Page 17; 119pp; English.
CC   The present sequence represents a cysteine rich soluble protein (CRSP)
CC   designated C10. CRSP proteins, and their (ant)agonists, are used to
CC   modulate physiology, differentiation, trafficking and development
CC   of cells (including those in culture), particularly cells of the
CC   mammalian immune system. They are used for treatment of abnormal
CC   proliferation (cancer, inflammation or degeneration), regeneration,
CC   degeneration and atrophy. The proteins are also used to raise, or
CC   detect, antibodies, to design oligonucleotides for library screening,
CC   in drug screens and to isolate cognate receptors. The antibodies are
CC   used for affinity purification of CRSP, to screen expression libraries,
CC   to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC   to produce anti-idiotypic antibodies (useful for diagnosis), in
CC   competitive drug screens, and as therapeutic modulators.
SQ   Sequence   111 AA;

Query Match        49.0%;     Score 284.5; DB 1; Length 111;
Best Local Similarity  49.0%;  Pred. No. 5.9e-21;
Matches   51; Conservative  21; Mismatches  31; Indels   1; Gaps   1;

Qy    4 LCLLLPVLGLLVSSKTLCSMEEAINERIQEVAGSLIFR-AISSIGLECQSVTSRGDLAT 62
         |  |:||:||| | |: ::|||||: :::|:|  :  | :: :|||:| | ::
Db    7 LLLILPLLQLINPGSTQCSLDSVMDKKIKDVLNSLEYSPSISKKLSCASVKSQGRPSS 66

Qy   63 CPRGFAVTGCTCGSACGSWDVRAETTCHCQCAGMDWTGARCCRV 106
         || | |||| ||  ||| |||||:||||||| : ||  |||:
Db   67 CPAGMAVTGCACGYGCGSWDVQLETTCHCQCSVVDWTTARCCHL 110

RESULT   6
W87706
ID   W87706 standard; Protein; 105 AA.
AC   W87706;
DT   09-MAR-1999 (first entry)
DE   A cysteine rich soluble protein designated C18.
KW   Cysteine rich soluble protein; CRSP; C18; cell development;
KW   mammalian immune system; antibody; abnormal proliferation; cancer;
KW   inflammation; degeneration; regeneration; degeneration; atrophy.
OS   Mus sp.
FH   Key         Location/Qualifiers
FT   Protein     1..19
FT               /note= "mature protein"
PN   WO9858061-A1.
PD   23-DEC-1998.
PF   18-JUN-1998; U12236.
PR   09-OCT-1997; US-061641.
PR   19-JUN-1997; US-878730.
PR   19-JUN-1997; US-878878.
PA   (SCHE ) SCHERING CORP.
PI   Franz-Bacon K, Gorman DM, McClanahan TK;
DR   WPI; 99-095339/08.
DR   N-PSDB; V84055.
PT   New cysteine-rich soluble proteins - used to modulate proliferation,
PT   differentiation, trafficking and development of cells, e.g. for
PT   treating inflammation, cancer and degeneration
PS   Claim 1; Page 14; 119pp; English.
CC   The present sequence represents a cysteine rich soluble protein (CRSP)
CC   designated C18. CRSP proteins, and their (ant)agonists, are used to
CC   modulate physiology, differentiation, trafficking and development
CC   of cells (including those in culture), particularly cells of the
CC   mammalian immune system. They are used for treatment of abnormal
CC   proliferation (cancer, inflammation or degeneration), regeneration,
CC   degeneration and atrophy. The proteins are also used to raise, or
CC   detect, antibodies, to design oligonucleotides for library screening,
CC   in drug screens and to isolate cognate receptors. The antibodies are
CC   used for affinity purification of CRSP, to screen expression libraries,
CC   to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC   to produce anti-idiotypic antibodies (useful for diagnosis), in
CC   competitive drug screens, and as therapeutic modulators.
SQ   Sequence   105 AA;

Query Match        47.1%;     Score 273.5; DB 1; Length 105;
Best Local Similarity  48.6%;  Pred. No. 6.5e-20;
Matches   51; Conservative  12; Mismatches  35; Indels   7; Gaps   2;

Qy    4 LCLL--LLPVLGLLVSSKTLCSMEEAINERIQEVAGSLIFRAISSIGLECQSVTSRGDLA 61
         | ||  |:: :|  |  ||  | ||  :::|:|  :||:| |:|:|||:| | ||
Db    5 LCFLFILVSLFPLIVPGNAQCSFESLVDQRIKEALSRQEPKTIS----CTSVTSSGRLA 59

Qy   62 TCPRGFAVTGCTCGSACGSWDVRAETTCHCQCAGMDWTGARCCRV 106
         :|| | |||| ||  ||| ||||||:||||||| : ||  |||:
Db   60 SCPAGMVTGCACGYGCGSWDIRNGNTCHCQCSVMDWASARCCRM 104
```

RESULT 7
W87705
ID   W87705 standard; Protein; 111 AA.
AC   W87705;
DT   09-MAR-1999 (first entry)
DE   A cysteine rich soluble protein designated C2b.
KW   Cysteine rich soluble protein; CRSP; C2b; cell development;
KW   mammalian immune system; antibody; abnormal proliferation; cancer;
KW   inflammation; degeneration; regeneration; atrophy.
OS   Mus sp.
FH   Key   Location/Qualifiers
FT   Protein   1..23
FT   /note= "mature protein"
PN   WO9858061-A1.
PD   23-DEC-1998.
PF   18-JUN-1998; U12236.
PR   09-OCT-1997; US-061641.
PR   19-JUN-1997; US-878730.
PR   19-JUN-1997; US-878878.
PA   (SCHE) SCHERING CORP.
PI   Franz-Bacon K, Gorman DM, McClanahan TK;
DR   WPI; 99-095339/08.
DR   N-PSDB; V84054.
PT   New cysteine-rich soluble proteins - used to modulate proliferation,
PT   differentiation, trafficking and development of cells, e.g. for
PT   treating inflammation, cancer and degeneration
PS   Claim 1; Page 13; 119pp; English.
CC   The present sequence represents a cysteine rich soluble protein (CRSP)
CC   designated C2b CRSP proteins, and their (ant)agonists, are used to
CC   modulate physiology, differentiation, trafficking and development
CC   of cells (including those in culture), particularly cells of the
CC   mammalian immune system. They are used for treatment of abnormal
CC   proliferation (cancer, inflammation or degeneration), regeneration,
CC   degeneration and atrophy. The proteins are also used to raise, or
CC   detect, antibodies, to design oligonucleotides for library screening,
CC   in drug screens and to isolate cognate receptors. The antibodies are
CC   used for affinity purification of CRSP, to screen expression libraries,
CC   to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC   to produce anti-idiotypic antibodies (useful for diagnosis), in
CC   competitive drug screens, and as therapeutic modulators.
SQ   Sequence   111 AA;

Query Match   42.3%; Score 245.5; DB 1; Length 111;
Best Local Similarity   40.2%; Pred. No. 3.7e-17;
Matches   43; Conservative   21; Mismatches   32; Indels   11; Gaps   2;

Qy   6 LLLLPVLGLLVSSKTLCSMEEAINERIQEVAG------SLIFRAISSIGLECQSVTSRGD 59
              |::::||:| :||: |:||||||||||: :::||| : |:|:|:||||||:: ||||
Db   9 LICISLLQLMVPVNTEGTLESIVEKKVKELLANRDDCPSTVTKTFS-----CTSITASGR 63

Qy   60 LATCPRGFAVTGCTCGSACGSWDVRAETTCHCQCAGMDWTGARCCRV 106
              ||:|| | |||| ||||||:|||||||:||::|||| |||||| :|
Db   64 LASCPSGMTVTGCACGYGCGSWDIRDGNTCHCQCSSTMDWATARCCQL 110

RESULT 8
Y12613
ID   Y12613 standard; Protein; 52 AA.
AC   Y12613;
DT   22-JUN-1999 (first entry)
DE   Human 5' EST secreted protein SEQ ID NO: 278 from WO 9906553.
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; antiinflammatory; tumour inhibition; antitumour.
OS   Homo sapiens.
PN   WO9906553-A2.
PD   11-FEB-1999.
PF   31-JUL-1998; IB1237,

PR   01-AUG-1997; US-905051.
PA   (GEST) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153783/13.
DR   N-PSDB; X41471.
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries derived from umbilical cord, lymph ganglia,
PT   lymphocytes and placental tissue
PS   Claim 34; Page 375; 41lpp; English.
CC   X41379 to X41526 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins, and encode the proteins given in Y12521 to Y12668,
CC   respectively. The proteins given represent the signal peptide and an
CC   N-terminal fragment of a secreted protein. The nucleic acid sequences
CC   can be used for producing human gene products. They can also
CC   be used to develop products for diagnosis and therapy. The proteins
CC   obtained may have cytokine activity, cell proliferation/differentiation
CC   activity, haematopoiesis regulating activity, tissue growth regulating
CC   activity, reproductive hormone regulating activity, chemotactic/
CC   chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC   ligand activity, antiinflammatory activity, tumour inhibition activity
CC   or other activities. The products can be used in forensic, gene therapy
CC   and chromosome mapping procedures. The sequences can also be used for
CC   obtaining corresponding promoter sequences. The nucleic acids encoding
CC   the signal peptide can be used for directing extracellular secretion of
CC   a polypeptide or the insertion of a polypeptide into a membrane, or
CC   importing a polypeptide into a cell.
SQ   Sequence   52 AA;

Query Match   39.1%; Score 227; DB 1; Length 52;
Best Local Similarity   100.0%; Pred. No. 1e-15;
Matches   48; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MKALCLLLPVLGLLVSSKTLCSMEEAINERIQEVAGSLIFRAISSIG 48
              |||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MKALCLLLPVLGLLVSSKTLCSMEEAINERIQEVAGSLIFRAISSIG 48

RESULT 9
W87704
ID   W87704 standard; Protein; 111 AA.
AC   W87704;
DT   09-MAR-1999 (first entry)
DE   A cysteine rich soluble protein designated C2.
KW   Cysteine rich soluble protein; CRSP; C2; cell development;
KW   mammalian immune system; antibody; abnormal proliferation; cancer;
KW   inflammation; degeneration; regeneration; degeneration; atrophy.
OS   Mus sp.
FH   Key   Location/Qualifiers
FT   Protein   1..23
FT   /note= "mature protein"
PN   WO9858061-A1.
PD   23-DEC-1998.
PF   18-JUN-1998; U12236.
PR   09-OCT-1997; US-061641.
PR   19-JUN-1997; US-878730.
PR   19-JUN-1997; US-878878.
PA   (SCHE) SCHERING CORP.
PI   Franz-Bacon K, Gorman DM, McClanahan TK;
DR   WPI; 99-095339/08.
DR   N-PSDB; V84053.
PT   New cysteine-rich soluble proteins - used to modulate proliferation,
PT   differentiation, trafficking and development of cells, e.g. for
PT   treating inflammation, cancer and degeneration
PS   Claim 1; Page 12-13; 119pp; English.
CC   The present sequence represents a cysteine rich soluble protein (CRSP)
CC   designated C2. CRSP proteins, and their (ant)agonists, are used to
CC   modulate physiology, differentiation, trafficking and development
CC   of cells (including those in culture), particularly cells of the
CC   mammalian immune system. They are used for treatment of abnormal
CC   proliferation (cancer, inflammation or degeneration), regeneration,
CC   degeneration and atrophy. The proteins are also used to raise, or
CC   detect, antibodies, to design oligonucleotides for library screening,

```
CC   in drug screens and to isolate cognate receptors. The antibodies are
CC   used for affinity purification of CRSP, to screen expression libraries,
CC   to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC   to produce anti-idiotypic antibodies (useful for diagnosis), in
CC   competitive drug screens, and as therapeutic modulators.
SQ   Sequence   111 AA;

Query Match             34.9%;  Score 202.5;  DB 1;  Length 111;
Best Local Similarity   36.3%;  Pred. No. 5.6e-13;
  Matches  37;  Conservative  24;  Mismatches 40;  Indels  1;  Gaps  1;

Qy    6 LLLLPVIGLLVSSKTLCSMEEAINERIQE-VAGSLIFRAISSIGLECQSVTSRGDLATCP 64
         |:||:|||| |  |:| | | :  :||  : :::||||| :|| :::|:| |
Db    9 LICISLLQLMVPVNTDETIEIIVENKVKELLANPANYPSTVTKTLSCTSVKTMNRWASCP 68

Qy   65 RGFAVTGCTCGSACGSWDVRAETTCHCQCAGMDWTGARCCRV 106
          | |||||||||||||||:::| :|||| ||||:|
Db   69 AGMTATGCACGFACGSWEIQSGDTCNCLCLLVDWTTARCCQL 110

RESULT 10
W68510
ID   W68510 standard; Protein; 1872 AA.
AC   W68510;
DT   06-JAN-1999   (first entry)
DE   Partial human Notch-3 protein.
KW   Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW   developmental cascade; neurogenic gene; mutant; neurological disorder;
KW   cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW   leukoencephalopathy.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 328
FT                   /note= "encoded by NAN"
FT   Misc_difference 401
FT                   /note= "encoded by GNN"
FT   Misc_difference 403
FT                   /note= "encoded by GNC"
FT   Misc_difference 406
FT                   /note= "encoded by GNN"
FT   Misc_difference 409
FT                   /note= "encoded by NNT"
FT   Misc_difference 420
FT                   /note= "encoded by GNC"
FT   Misc_difference 706
FT                   /note= "encoded by NNN"
FT   Misc_difference 708
FT                   /note= "encoded by CCN"
FT   Misc_difference 719
FT                   /note= "encoded by CGN"
FT   Misc_difference 728
FT                   /note= "encoded by CNT"
FT   Misc_difference 729
FT                   /note= "encoded by GTN"
FT   Misc_difference 759..789
FT                   /note= "encoded by NNN"
FT   Misc_difference 1425
FT                   /note= "encoded by GNA"
PN   FR2751985-A1.
PD   06-FEB-1998.
PF   01-AUG-1996;  009733.
PR   01-AUG-1996;  FR-009733.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
     WPI; 98-133137/13.
DR   N-PSDB; V57163.
PT   Human Notch3 nucleic acids - and methods for identifying
PT   pre-disposition to cerebral autosomal dominant arteriopathy with
PT   sub-cortical infarcts and leukoencephalopathy
PS   Claim 2; Fig 1a-1g; 42pp; French.
CC   This sequence represents a partial human notch3 protein, a transmembrane
CC   receptor protein involved in lateral inhibition and regulating
```

```
CC   developmental cascades of neurogenic genes.  Mutated Notch3 proteins
CC   are thought to be involved in neurological disorders, especially of the
CC   cerebral autosomal dominant arteriopathy with subcortical infarcts and
CC   leukoencephalopathy (CADASIL) type.  Blocking expression of a mutated
CC   Notch3 gene or by substitution therapy with non-mutated Notch3 gene or
CC   protein can be used to treat CADASIL or related disorders.
SQ   Sequence   1872 AA;

Query Match             14.1%;  Score 82;  DB 1;  Length 1872;
Best Local Similarity   29.9%;  Pred. No. 6.6;
  Matches  23;  Conservative   7;  Mismatches 23;  Indels 24;  Gaps  4;

Qy   47 IGLECQSVTSRGDLATCPRGFAVTGCTCGSACGSWDVRAETTC------------- 90
          :|: || ||| | ||:| ||| | :| | |  :| :|
Db 1234 VGVPCQQ-TPRGPRCACPPG--LSGPSCRSFPGSPPGASNASCAAAPCLHGGSCRPAPLA 1290

Qy   90 ---HCQCAGMDWTGARC 103
           || ||| ||
Db 1291 PFFRCACA-QGWTGPRC 1306

RESULT 11
W49698
ID   W49698 standard; Protein; 2321 AA.
AC   W49698;
DT   21-DEC-1998   (first entry)
DE   Human Notch3 protein.
KW   Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW   developmental cascade; neurogenic gene; mutant; neurological disorder;
KW   cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW   leukoencephalopathy; therapy.
OS   Homo sapiens.
PN   FR2751986-A1.
PD   06-FEB-1998.
PF   16-APR-1997;  004680.
PR   01-AUG-1996;  FR-009733.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Bach JF, Bousser MG, Joutel A, Tournier Lasserve E;
     WPI; 98-133138/13.
DR   N-PSDB; V57001.
PT   Human Notch3 nucleic acids - and methods for identifying
PT   pre-disposition to cerebral autosomal dominant arteriopathy with
PT   sub-cortical infarcts and leukoencephalopathy
PS   Claim 2; Fig 1.1-1.8; 45pp; French.
CC   This sequence represents the human Notch3 protein, a transmembrane
CC   receptor protein involved in lateral inhibition and regulating
CC   developmental cascades of neurogenic genes.  Mutated Notch3 proteins
CC   are thought to be involved in neurological disorders, especially of
CC   the cerebral autosomal dominant arteriopathy with subcortical infarcts
CC   and leukoencephalopathy (CADASIL) type.  Blocking expression of a
CC   mutated Notch3 gene or by substitution therapy with non-mutated Notch3
CC   gene or protein can be used to treat CADASIL or related disorders.
SQ   Sequence   2321 AA;

Query Match             14.1%;  Score 82;  DB 1;  Length 2321;
Best Local Similarity   29.9%;  Pred. No. 8.4;
  Matches  23;  Conservative   7;  Mismatches 23;  Indels 24;  Gaps  4;

Qy   47 IGLECQSVTSRGDLATCPRGFAVTGCTCGSACGSWDVRAETTC------------- 90
          :|: || ||| | ||:| ||| | :| | |  :| :|
Db 1300 VGVPCQQ-TPRGPRCACPPG--LSGPSCRSFPGSPPGASNASCAAAPCLHGGSCRPAPLA 1356

Qy   90 ---HCQCAGMDWTGARC 103
           || ||| ||
Db 1357 PFFRCACA-QGWTGPRC 1372

RESULT 12
R98207
ID   R98207 standard; Protein; 60 AA.
AC   R98207;
```

DT   30-DEC-1996   (first entry)
DE   Nucleotide used in production of MSH/MoMuLV chimeric sequence.
KW   Moloney murine leukaemia virus; gp70; 4070A retrovirus; retrovirus;
KW   10A1 murine leukaemia virus; NZB-9-1 murine leukaemia virus;
KW   polytropic MX27 provirus; targeted drug delivery; gene therapy;
KW   single chain antibody; envelope protein; ss.
OS   Synthetic.
PN   WO9630504-A1.
PD   03-OCT-1996.
PF   22-MAR-1996; U03908.
PR   24-MAR-1995; US-409648.
PA   (GENE-) GENETIC THERAPY INC.
PA   (UYSC-) UNIV SOUTHERN CALIFORNIA.
PI   Anderson W, Chiang YL, Januszeski M, Mackrell AJ;
PI   Zhao Y;
DR   WPI: 96-455352/45.
PT   Cell-targetted retroviral vector particles - having envelope protein
PT   modified with targetting polypeptide
PS   Example 2; Page 36; 73pp; English.
CC   Cell targetted retroviral vector particles can be used in gene
CC   therapy to deliver a heterologous gene to a target cell. The cell
CC   expression of a heterologous polypeptide in that cell. The cell
CC   targeted retroviral vector particles comprise an envelope protein
CC   which is modified to contain a targetting polypeptide (a single chain
CC   antibody), or in the case of moloney murine leukaemia virus
CC   (MoMuLV), alpha melanotropin-stimulating hormone (MSH). Two
CC   oligonucleotides (R98207, R98208) were used to substitute sequences in
CC   MoMuLV for MSH sequences. This oligonucleotide was used to replace
CC   residues S74-591 of MoMuLV envelope protein (See W04248).
SQ   Sequence 60 AA;

Query Match   13.2%; Score 76.5; DB 1; Length 60;
Best Local Similarity   35.1%; Pred. No. 0.51;
Matches 20; Conservative 3; Mismatches 19; Indels 15; Gaps 3;

Qy   48 GLECQSVTSRGDLATCPRGFAVTGCTCGSACGSWDVRAETTCHCQCAGMDWTGARCC 104
          | ||| | || |||| |||||||| || |||||||| || ||
Db   19 GMACATTTTCG--ATGGGTAAGGCCCTCAC--------CCCTCGG---TGCAAC 60

RESULT 13
W74444
ID   W74444 standard; Protein; 131 AA.
AC   W74444;
DT   11-MAY-1999   (first entry)
DE   Mouse Tie receptor tyrosine kinase fragment.
KW   Tie gene; receptor tyrosine kinase; promoter; gene expression;
KW   human gene therapy; growth factor.
OS   Mus sp.
PN   US5877020-A.
PD   02-MAR-1999.
PF   31-MAY-1996; 650598.
PR   31-MAY-1996; US-650598.
PR   22-SEP-1994; US-310717.
PA   (UYHE-) UNIV HELSINKI LICENSING LTD.
PI   Alitalo K;
DR   WPI: 99-189653/16.
PT   Tie receptor tyrosine kinase promoter - for directing expression of
PT   recombinant DNA in endothelial cells
PS   Example 1; Column 6; 26pp; English.
CC   This sequence represents a fragment of the mouse Tie receptor tyrosine
CC   kinase. The invention relates to Tie receptor tyrosine kinase promoters.
CC   The promoters are useful for directing expression of recombinant DNA
CC   sequences in endothelial cells. The promoters are useful for production
CC   of proteins and peptides which act as anticoagulants, vasodilator
CC   inhibitors of thrombosis or restenosis into endothelial cells, blood and
CC   tissues. The promoters are useful for directing expression of proteins
CC   and peptides for human gene therapy, antigens and markers for endothelial
CC   cell tagging, and antisense RNA constructs for use in endothelial cells
CC   in vivo and in vitro. The promoters, and vectors and host cells
CC   containing them, are useful in gene therapy for promoting expression of
CC   various growth factors or receptors or their domains.

SQ   Sequence 131 AA;

Query Match   12.6%; Score 73; DB 1; Length 131;
Best Local Similarity   28.8%; Pred. No. 2.6;
Matches 23; Conservative 4; Mismatches 19; Indels 34; Gaps 5;

Qy   59 DLATCPRGFAVTGCTCGSAC-----------GSWDVRAETTC----HCQCAGMDWT 99
          | | || ||| || | |||
Db   19 DGCVCPPGF--TGTRCEQACREGRFGQSCQEQCPGTAGCRGLTFCLPPYGCSC-GSGWR 75

Qy   100 GARC------------CRVQ 107
           ||| |||
Db   76 GSQCQEACAPDHFGADCRLQ 95

RESULT 14
R91428
ID   R91428 standard; Protein; 1111 AA.
AC   R91428;
DT   13-NOV-1996   (first entry)
DE   Kalinin/laminin 5 gamma-2 chain (alternative form)
KW   kalinin; laminin; epidermolysis bullosa; junctional; probe;
KW   detection; inhibit; monitor; malignancy.
OS   Homo sapiens.
PN   WO9610646-A1.
PD   11-APR-1996.
PF   04-OCT-1995; E03918.
PR   04-OCT-1994; US-317450.
PA   (TRYG/) TRYGGVASON K.
PI   Kallunki P, Pyke C, Tryggvason K;
DR   N-PSDB; T13323.
DR   WPI; 96-209366/21.
PT   Detection of kalinin or laminin 5 expression in cells - useful to
PT   detect, monitor and inhibit the invasive growth of cell in tissue,
PT   partic. malignant tissue
PS   Disclosure; Fig 4B; 37pp; English.
CC   The present sequence is an alternative form of kalinin/laminin 5 gamma-2
CC   chain (see R91427). The gamma-2 chain is of importance to patients
CC   suffering from epidermolysis bullosa, esp. the junctional form (JEB).
CC   Probes and antisense gamma-2 sequences derived from this sequence can be
CC   used to detect, monitor and inhibit the invasive growth of cells in
CC   tissue, partic. malignant tissue.
SQ   Sequence 1111 AA;

Query Match   12.4%; Score 72; DB 1; Length 1111;
Best Local Similarity   25.4%; Pred. No. 35;
Matches 31; Conservative 14; Mismatches 47; Indels 30; Gaps 8;

Qy   4 LCLLLLPVLGLLVSSKTLC-----SMEEAINERIQEVAGSLIFRAI---SSIGLECQSV 54
         || || | ||| : |: |: :|: |||| : ||| |
Db   10 LCFSLLLPAARATSRREVCDCNGKSRQCIFDRELHRQTGN-GFRCLNCDNTDGIHCEK- 68

Qy   55 TSRGDLATCPRGF-------AVTGCTCGSACGSWDVRAETTCHCQC-AGMDWTGARCCRV 106
          ||| |||||||| : ||||||| ||| |||||
Db   68 --------CKNGFYRHRERDRCLPCNCNSK-GSLSARCDNSGRCSCKPGV--TGARCDRC 116

Qy   107 QP 108
           |
Db   117 LP 118

RESULT 15
R91427
ID   R91427 standard; Protein; 1193 AA.
AC   R91427;
DT   13-NOV-1996   (first entry)
DE   Kalinin/laminin 5 gamma-2 chain.
KW   kalinin; laminin; epidermolysis bullosa; junctional; probe;
KW   detection; inhibit; monitor; malignancy.
OS   Homo sapiens.
PN   WO9610646-A1.

Fri May 26 10:31:08 2000

```
PD    11-APR-1996.
PF    04-OCT-1995; E03918.
PR    04-OCT-1994; US-317450.
PA    (TRYG/) TRYGGVASON K.
PI    Kallunki P,  Pyke C,  Tryggvason K;
DR    WPI: 96-209366/21.
DR    N-PSDB; T13323.
PT    Detection of kalinin or laminin 5 expression in cells - useful to
PT    detect, monitor and inhibit the invasive growth of cell in tissue,
PT    partic. malignant tissue
PS    Disclosure; Fig 4A; 37pp; English.
CC    The present sequence is the kalinin/laminin 5 gamma-2 chain. The gamma-2
CC    chain is of importance to patients suffering from epidermolysis bullosa,
CC    esp. the junctional form (JEB). Probes and antisense gamma-2 sequences
CC    derived from this sequence can be used to detect, monitor and inhibit
CC    the invasive growth of cells in tissue, partic. malignant tissue.
SQ    Sequence  1193 AA;


Query Match             12.4%;   Score 72;  DB 1;  Length 1193;
Best Local Similarity   25.4%;   Pred. No. 38;
Matches   31;  Conservative  14;  Mismatches  47;  Indels  30;  Gaps  8;

Qy    4 LCLLLPVLGLLVSSKTLC-----SMEEAINERIQEVAGSLIFRAI----SSIGLECQSV 54
        || || ||  | :: :|     | ||  | : ||  ::  | ::|:: ||
Db   10 LCFSLLLPAARATSRREVCDCNGKSRQCIFDRELHRQTGN-GFRCLNCNDNTDGIHCEK- 68

Qy   55 TSRGDLATCPRGF-------AVTGCTCGSACGSWDVRAETTCHCQC-AGMDWTGARCCRV 106
         : ||       |       | | | ||| | :: : | |||| : |  |||||
Db   68 ------CKNGFYRHRERDRCLPCNCNSK-GSLSARCDNSGRCSCKPGV--TGARCDRC 116

Qy  107 QP 108
         |
Db  117 LP 118


Search completed: May 26, 2000, 03:02:14
Job time: 10557 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:   May 26, 2000, 00:05:24 ; Search time 51.41 Seconds
                                   (without alignments)
                                   2204.571 Million cell updates/sec

Title:          US-09-099-898-1
Perfect score:  453
Sequence:       1 GTGTGCCGGATTGGTTAGC.........GGGGAGCTGGAAATAAACCT 453

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters:    623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :   N_Geneseq_36:*

  Pred. No. is the number of results predicted by chance to have a
  score greater than or equal to the score of the result being printed,
  and is derived by analysis of the total score distribution.

                          SUMMARIES

                  %
Result        Query
No.    Score  Match  Length DB  ID        Description
-----  -----  -----  ------ --  --        -----------
  1     453   100.0    453   1  V84059    cDNA encoding a cy
  2    420.6   92.8    553   1  X51720    DNA encoding a hum
  3    194     42.8    229   1  X41471    Human secreted pro
  4    158.4   35.0    572   1  V84057    cDNA encoding a cy
  5    152.4   33.6    560   1  V84056    cDNA encoding a cy
  6     92.8   20.5    603   1  V84058    cDNA encoding a cy
  7     78     17.2    554   1  V84055    cDNA encoding a cy
  8     64.4   14.2    574   1  V84054    cDNA encoding a cy
  9     54.4   12.0    527   1  V84053    cDNA encoding a cy
 10     40.6    9.0    985   1  V44439    Mycobacterium tube
 11     40.6    9.0    985   1  V64548    M. tuberculosis im
 12     39      8.6   8438   1  Q73500    DNA encoding Pseud
 13     38.6    8.5 114955   1  X53491    Human integrin Al
 14     36.6    8.1   3415   1  T36481    Human integrin bet
 15     36.6    8.1  11820   1  V18130    Human chromosome 1
 16     36.2    8.0   2932   1  Q25388    TXA2 receptor gene
 17     35.2    7.8    329   1  V44425    Mycobacterium tube
 18     35.2    7.8    329   1  V64534    M. tuberculosis im
 19     34.8    7.7  30001   1  T61016    Total DNA sequence
 20     34.8    7.7  30001   1  X05110    S. aureofaciens DN
 21     34.8    7.7 114955   1  X53491    Human adenosine Al
 22     34.4    7.7 117213   1  V62176    HSV-2 strain SB5 C
 23     34.4    7.6    201   1  N70195    Streptomyces prote
 24     34.2    7.5   1035   1  T49316    Hypersensitive res
 25     34.2    7.5   1035   1  X36430    Hypersensitive res
 26     34.2    7.5   1035   1  V39975    Pseudomonas solana
 27     34.2    7.5   1035   1  V54609    Pseudomonas solana
 28     34.2    7.5   1035   1  V83991    DNA encoding a hyp
 29     34      7.5    201   1  X19063    Human PPAR-gamma-1
 30     34      7.5    384   1  V23147    Rat Pualpha like
 31     34      7.5   1545   1  V41906    Nucleotide sequenc
 32     34      7.5   1687   1  T39798    Human clone 56 gen
 33     33.8    7.5    535   1  V44428    Mycobacterium tube
 34     33.8    7.5    535   1  V64537    M. tuberculosis im
 35     33.8    7.5   6854   1  T76903    S. glaucescens Pst
 36     33.6    7.4   1236   1  V72041    Human CPG2/R6 vari
 37     33.6    7.4   1266   1  X09011    Brn-3a polynucleot
 38     33.6    7.4   1439   1  V44424    Mycobacterium tube
 39     33.6    7.4   1439   1  V64533    M. tuberculosis im
 40     33.6    7.4   1926   1  V72054    806.077 heavy chai
 41     33.6    7.4   1929   1  V72046    Clone pNG4/A5B7VH-
 42     33.6    7.4   1998   1  V72075    Fusion protein (80
 43     33.6    7.4   2019   1  V72059    Plasmid pNG4/55.1s
 44     33.6    7.4   2025   1  V72064    Fusion protein pNG
 45     33.6    7.4   3217   1  V72076    IRES-based A5B7 ch
```

ALIGNMENTS

```
RESULT   1
V84059
ID   V84059 standard; cDNA; 453 BP.
AC   V84059;
DT   09-MAR-1999  (first entry)
DE   cDNA encoding a cysteine rich soluble protein designated C23.
KW   Cysteine rich soluble protein; CRSP; C23; cell development;
KW   mammalian immune system; antibody; abnormal proliferation; cancer;
KW   inflammation; degeneration; regeneration; degeneration; atrophy; ss.
OS   Homo sapiens.
FH   Key            Location/Qualifiers
FT   CDS            47..373
FT                  /*tag= a
FT                  /product= C23
FT   sig_peptide    47..100
FT                  /*tag= b
FT   mat_peptide    101..370
FT                  /*tag= c
PN   WO9858061-A1.
PD   23-DEC-1998.
PF   18-JUN-1998;  U12236.
PR   09-OCT-1997;  US-061641.
PR   19-JUN-1997;  US-878730.
PR   19-JUN-1997;  US-878878.
PA   (SCHE ) SCHERING CORP.
PI   Franz-Bacon K, Gorman DM, McClanahan TK;
DR   WPI; 99-095339/08.
DR   P-PSDB; W87710.
PT   New cysteine-rich soluble proteins - used to modulate proliferation,
PT   differentiation, trafficking and development of cells, e.g. for
PT   treating inflammation, cancer and degeneration
PS   Claim 16; Page 18-19; 119pp; English.
CC   The present sequence encodes a cysteine rich soluble protein (CRSP)
CC   designated C23. CRSP proteins, and their (antagonists), are used to
CC   modulate physiology, differentiation, trafficking and development
CC   of cells (including those in culture), particularly cells of the
CC   mammalian immune system. They are used for treatment of abnormal
CC   proliferation (cancer, inflammation or degeneration), regeneration,
CC   degeneration and atrophy. The proteins are also used to raise, or
CC   detect, antibodies, to design oligonucleotides for library screening,
CC   in drug screens and to isolate cognate receptors. The antibodies are
CC   used for affinity purification of CRSP, to screen expression libraries,
CC   to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC   to produce anti-idiotypic antibodies (useful for diagnosis), in
CC   competitive drug screens, and as therapeutic modulators.
SQ   Sequence    453 BP;     74 A;    134 C;    158 G;     87 T;

Query Match             100.0%;  Score 453;  DB 1;  Length 453;
Best Local Similarity   100.0%;  Pred. No. 1.7e-106;
Matches  453;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 GTGTGCCGGATTGGTTAGCTGAGCCACCGAGAGCGCCCTGCAGGATGAAAGCTCTG 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTGTGCCGGATTGGTTAGCTGAGCCACCGAGAGCGCCCTGCAGGATGAAAGCTCTG 60

Qy   61 TCTCCTCCTCCTCCCGTCTCCCGGGCTGTTGGTCTTCTAGCAAGACCCTGTCTCCATGGA 120
```

Right-column continuation:

```
CC   disorders, skeletal disorders, neurological disorders, arthritic
CC   disorders, asthma, immunodeficiency diseases, AIDS and transplant
CC   rejection. The polypeptides are also useful for identifying their
CC   binding partners.
SQ   Sequence 553 BP;  160 A;  132 C;  167 G;   94 T;

Query Match             92.8%;  Score 420.6;  DB 1;  Length 553;
Best Local Similarity   98.5%;  Pred. No. 3e-98;
Matches   446;  Conservative   0;  Mismatches   4;  Indels   3;  Gaps   2;

Qy     1 GTGTGCGGATTGGTTAGCTGAGCCGCCAGGAGGGCCGCTGCAGGATGAAAGCTCTCTG  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GTGTGCGGATTGGTTAGCTGAGCCGCCAGGAGGGCCGCTGCAGGATGAAAGCTCTCTG  60

Qy    61 TCTCCTCCTCCTCCGTCCGGCTCCCAGGGGGACCTGGCTCCGTGTCAAGACCCGTGTCCATGGA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TCTCCTCCTCCTCCGTCCGGCTCCCAGGGGGACCTGGCTCCGTGTCAAGACCCGTGTCCATGGA 120

Qy   121 AGAAGCCATCAATGAGAGAGGATCCAGAGAGGTCGCCGGCTCCTAATATTAGGCAATAAG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGAAGCCATCAATGAGAGAGGATCCAGAGAGGTCGCCGGCTCCTAATATTAGGCAATAAG 180

Qy   181 CAGCATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTACTTGCCCCCG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CAGCATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTACTTGCCCCCG 240

Qy   241 AGGCTTCGCCGTCACCGGCTGCACTTGTGGCTCCGCCCTGGCTCGTGGGATGTGCGCGC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AGGCTTCGCCGTCACCGGCTGCACTTGTGGCTCCGCCCTGGCTCGTGGGATGTGCGCGC 300

Qy   301 CGAGACCACATGTCACTGCCAGTGGACTGGACGCGGCGAGGCGGCTCCGTGTCG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CGAGACCACATGTCACTGCCAGTGGACTGGACGCGGCGAGGCGGCTCCGTGTCG 360

Qy   361 TGTGCAGCCCTGAGGTCGCGGCGAGCGGCGTGCACAGCGGGCGGGAGGCGGCTCCAGGTC 420
           ||||||||||||||||||||||  |  |||||||||||||||||||||||||
Db   361 TGTGCAGCCCTGAGGTCGCGGCCA--GTGGCAACAGCGGCGGGGCGGCTCCAGGTC 418

Qy   421 CGGAGGGGTTGCGGGGGAGCTGGAAATAAACCT 453
           ||||  |||||||||||||||||||||||||
Db   419 CGGA-GGGTTCGGGGGGAGCTGGAAATAAACCT 450

RESULT  3
X41471
ID   X41471  standard; cDNA; 229 BP.
AC   X41471;
DT   22-JUN-1999  (first entry)
DE   Human secreted protein 5' EST SEQ ID NO: 130 from WO 9906553.
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; antiinflammatory; tumour inhibition; antitumour; ds.
OS   Homo sapiens.
PN   WO9906553-A2.
PD   11-FEB-1999.
PF   31-JUL-1998;  IB1237.
PR   01-AUG-1997;  US-905051.
PA   (GEST ) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153783/13.
DR   P-PSDB; Y12613.
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries derived from umbilical cord, lymph ganglia,
PT   lymphocytes and placental tissue
PS   Claim 1; Page 238; 41pp; English.
CC   X41379 to X41526 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins, and encode the proteins given in Y12521 to Y12668
CC   respectively. The proteins given represent the signal peptide and an
CC   N-terminal fragment of a secreted protein. The nucleic acid sequences
```

Left-column content:

```
           |||||||||||||||||||||||||||||||||||||||||
Db    61 TCTCCTCCTCCGTCCGGCTCCCAGGGGGACCTGGCTACTTGCCCCCG

Qy   121 AGAAGCCATCAATGAGAGAGGATCCAGAGAGGTCGCCGAGCTCCTAATATTAGGCAATGA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGAAGCCATCAATGAGAGAGGATCCAGAGAGGTCGCCGAGCTCCTAATATTAGGCAATGA 180

Qy   181 CAGCATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTACTTGCCCCCG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CAGCATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTACTTGCCCCCG 240

Qy   241 AGGCTTCGCCGTCACCGGCTGCACTTGTGGCTCCGCCCTGGCTCGTGGGATGTGCGGC 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AGGCTTCGCCGTCACCGGCTGCACTTGTGGCTCCGCCCTGGCTCGTGGGATGTGCGGC 300

Qy   301 CGAGACCACATGTCACTGCCAGTGGACTGGACGCGGCGAGGCGGCTCCGTGTCG 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CGAGACCACATGTCACTGCCAGTGGACTGGACGCGGCGAGGCGGCTCCGTGTCG 360

Qy   361 TGTGCAGCCCTGAGGTCGCGGCGAGCGGCGTGCACAGCGGGCGGGAGGCGGCTCCAGGTC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGTGCAGCCCTGAGGTCGCGGCGAGCGGCGTGCACAGCGGGCGGGAGGCGGCTCCAGGTC 420

Qy   421 CGGAGGGGTTGCGGGGGAGCTGGAAATAAACCT 453
           |||||||||||||||||||||||||||||||||
Db   421 CGGAGGGGTTGCGGGGGAGCTGGAAATAAACCT 453

RESULT  2
X51720
ID   X51720  standard; DNA; 553 BP.
AC   X51720;
DT   17-JUN-1999  (first entry)
DE   DNA encoding a human secreted protein.
KW   Human secreted protein; cancer; immune disorder; infection;
KW   inflammatory disorder; skin disorder; tumour; atherosclerosis;
KW   restenosis; autoimmune disorder; Alzheimer's disease;
KW   peripheral neuropathy; trauma; spinal cord injury; allergy;
KW   hematopoietic disorder; skeletal disorder; neurological disorder;
KW   arthritic disorder; asthma; immunodeficiency disease; AIDS;
KW   transplant rejection; ss.
OS   Homo sapiens.
PN   WO9911293-A1.
PD   11-MAR-1999.
PF   03-SEP-1998;  U18360.
PR   12-SEP-1997;  US-058974.
PR   05-SEP-1997;  US-057526.
PR   05-SEP-1997;  US-057663.
PR   05-SEP-1997;  US-057669.
PR   12-SEP-1997;  US-058666.
PR   12-SEP-1997;  US-058667.
PR   12-SEP-1997;  US-058973.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Brewer LA, Ebner R, Lafleur DW, Moore PA, Olsen HS,
PI   Rosen GA, Ruben SM, Shi Y;
DR   WPI; 99-204988/17.
DR   P-PSDB; Y12933.
PT   New isolated human genes and the secreted polypeptides they encode
PT   - useful for diagnosis and treatment of e.g. neurological disorders,
PT   tumours, immune disorders, inflammation or haematological disorders
PS   Claim 1; Page 165; 215pp; English.
CC   X51701-55 encode human secreted proteins. The polynucleotides and
CC   their corresponding secreted polypeptides are useful for preventing,
CC   treating or ameliorating medical conditions, e.g. by protein or gene
CC   therapy. Pathological conditions can also be diagnosed by determining
CC   the amount of the new polypeptides in a sample or by determining the
CC   presence of mutations in the new polynucleotides. Specific uses are
CC   described for each polynucleotide, based on which tissues they are
CC   most highly expressed in, and include developing products for the
CC   diagnosis or treatment of cancer, immune disorders, infection,
CC   inflammatory disorders, skin disorders, tumours, atherosclerosis,
CC   restenosis, autoimmune disorders, Alzheimer's disease, peripheral
CC   neuropathies, trauma, spinal cord injuries, allergy, hematopoietic
```

CC   can be used for producing secreted human gene products. They can also
CC   be used to develop products for diagnosis and therapy. The proteins
CC   obtained may have cytokine activity, cell proliferation/differentiation
CC   activity, haematopoiesis regulating activity, tissue growth regulating
CC   activity, reproductive hormone regulating activity, chemotactic/
CC   chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC   ligand activity, antiinflammatory activity, tumour inhibition activity
CC   or other activities. The products can be used in forensic, gene therapy
CC   and chromosome mapping procedures. The sequences can also be used for
CC   obtaining corresponding promoter sequences. The nucleic acids encoding
CC   the signal peptide can be used for directing extracellular secretion of
CC   a polypeptide or the insertion of a polypeptide into a membrane, or
CC   importing a polypeptide into a cell.
SQ   Sequence 229 BP; 51 A; 58 C; 71 G; 48 T;

Query Match   42.8%; Score 194; DB 1; Length 229;
Best Local Similarity 97.5%; Pred. No. 7.4e-41;
Matches 197; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy   1 GTGTGCCGGATTTGGTTAGCTGAGCCCACCGAGAGGCGCCTGCAGGATGAAAGCTCTCTG 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   28 GTGTGCCGGATTTGGTTAGCTGAGCCCACCGAGAGGCGCCTGCAGGATGAAAGCTCTCTG 87

Qy   61 TCTCCTCCTCCTCCCTCCTCGGGGCTGTTGGTGTCTAGCAAGACCCTGTCTCCATGGA 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   88 TCTCCTCCTCCTCCCTCCTCGGGGCTGTTGGTGTCTAGCAAGACCCTGTCTCCATGGA 147

Qy 121 AGAAGCCATCAATGAGAGGATCCAGGAGGTCGCCGCCTCCCTAATATTTAGGGCAATAAG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 148 AGAAGCCATCAATGAGAGGATCCAGGAGGTCGCCGCCTCCCTAATATTTAGGGCAATAAG 207

Qy 181 CAGCATTGGCCTGGAGTGCCAG 202
         |||||||||| |-|||- ||
-Db 208 CAGCATTGGCCGAGGGAGCCAG 229

RESULT 4
ID   V84057 standard; cDNA; 572 BP.
AC   V84057;
DT   09-MAR-1999 (first entry)
DE   cDNA encoding a cysteine rich soluble protein designated C19.
KW   Cysteine rich soluble protein; CRSP; C19; cell development;
KW   mammalian immune system; antibody; abnormal proliferation; cancer;
KW   inflammation; degeneration; regeneration; degeneration; atrophy; ss.
OS   Rattus sp.
FH   Key   Location/Qualifiers
FT   CDS   29..373
FT   /*tag= a
FT   /product= C19
FT   sig_peptide   29..79
FT   /*tag= b
FT   mat_peptide   80..370
FT   /*tag= c
FT   misc_feature   159..160
FT   /*tag= d
FT   /note= "intron present between these nucleotides in
FT   the genomic DNA"
FT   misc_feature   236..237
FT   /*tag= e
FT   /note= "intron present between these nucleotides in
FT   the genomic DNA"
PN   WO9858061-A1.
PD   23-DEC-1998.
PF   18-JUN-1998; U12236.
PR   09-OCT-1997; US-061641.
PR   19-JUN-1997; US-878730.
PR   19-JUN-1997; US-878878.
PA   (SCHE) SCHERING CORP.
PI   Franz-Bacon K, Gorman DM, McClanahan TK;
DR   WPI; 99-095339/08.
DR   P-PSDB; W87708.

PT   New cysteine-rich soluble proteins - used to modulate proliferation,
PT   differentiation, trafficking and development of cells, e.g. for
PT   treating inflammation, cancer and degeneration
PS   Claim 16; Page 16; 119pp; English
CC   The present sequence encodes a cysteine rich soluble protein (CRSP)
CC   designated C19. CRSP proteins, and their (ant)agonists, are used to
CC   modulate physiology, differentiation, trafficking and development
CC   of cells (including those in culture), particularly cells of the
CC   mammalian immune system. They are used for treatment of abnormal
CC   proliferation (cancer, inflammation or degeneration), regeneration,
CC   degeneration and atrophy. The proteins are also used to raise, or
CC   detect, antibodies, to design oligonucleotides for library screening,
CC   in drug screens and to isolate cognate receptors. The antibodies are
CC   used for affinity purification of CRSP, to screen expression libraries,
CC   to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC   to produce anti-idiotypic antibodies (useful for diagnosis), in
CC   competitive drug screens, and as therapeutic modulators.
SQ   Sequence 572 BP; 141 A; 141 C; 166 G; 124 T;

Query Match   35.0%; Score 158.4; DB 1; Length 572;
Best Local Similarity 69.2%; Pred. No. 9.4e-32;
Matches 216; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

Qy   56 CTCTGTCTCCTCCTCCTCCCTCCTGGGGCTGTTGGTGTCTAGCAAGACCCTGTCTGC 115
         ||| |||||||||||||||||||||| | ||||||| | - |||||-|||||||
Db   50 CTCCTTTCCTTTCTTCCTTGTCTGGGGCCTGTGGGCCCCAGCATGTCACTGTGTCCC 109

Qy 116 ATGGAAGAAGCCATCAATGAGAGGATCCAGGAGGTCGCCGCCTCCCTAATATTTAGGGCA 175
         ||||| ||||||| |||| |||| |||| || |-|||||||||
Db 110 ATGATGAAGCCATCAGCAAGAAGATCAATCAAGACTTCAGCTCCTCCTACTGCCAGCTGCA 169

Qy 176 ATAAGCAGCATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGGACCTGGCTACTTGC 235
         || |||| ||||-|| ||| ||||||||| |||||||| |||||
Db 170 ATGAAGAACAACTGCCTACATTGGCTGGTCAGTCTCCTCCAGAGGGAGGCTGCCCCCTGC 229

Qy 236 CCCGGAGGCTTCGGCCGTCACCGGCTGCACTTGTGGCTCCGCTCGTGGCTCCGTCGTGGATGTG 295
         || |||| |||||||| |||||||| ||||||| ||| ||||||
Db 230 CCAGAAGGCACAACCGTCACTAGCTGCTCACTGGCTCCTGGCTCCGTGGCTCATGGACGTC 289

Qy 296 CGCCCGAGCCACCATGTCACTGCCAGTGCGGCGGCATGGACTGACGCCGAGCGCCGCTGC 355
         || |||| | ||| |||||||| ||| |||||| ||||||
Db 290 CGTGGAGGATACAATGTCACTGCCAGTGCGGCGAAGCATAGACTGGACAGCGCCCCGGCTGC 349

Qy 356 TGTCGTGCGCAG 367
         ||| |||||
Db 350 TGTACCCTGCGG 361

RESULT 5
ID   V84056 standard; cDNA; 560 BP.
AC   V84056;
DT   09-MAR-1999 (first entry)
DE   cDNA encoding a cysteine rich soluble protein designated C19.
KW   Cysteine rich soluble protein; CRSP; C19; cell development;
KW   mammalian immune system; antibody; abnormal proliferation; cancer;
KW   inflammation; degeneration; regeneration; degeneration; ss.
OS   Mus sp.
FH   Key   Location/Qualifiers
FT   CDS   64..408
FT   /*tag= a
FT   /product= C19
FT   sig_peptide   64..123
FT   /*tag= b
FT   mat_peptide   124..405
FT   /*tag= c
FT   misc_feature   193..194
FT   /*tag= d
FT   /note= "intron present between these nucleotides in
FT   the genomic DNA"
FT   misc_feature   271..272
FT   /*tag= e

```
FT                   /note= "intron present between these nucleotides in
FT                          the genomic DNA"
FT   sig_peptide     108..167
FT                   /tag= a
FT                   /product= C10
FT   mat_peptide     168..440
FT                   /tag= b
FT   misc_feature    234..235
FT                   /tag= c
FT                   /note= "introns present between these nucleotides
FT                          in the genomic DNA"
FT   misc_feature    315..316
FT                   /tag= e
FT                   /note= "introns present between these nucleotides
FT                          in the genomic DNA"
PN   WO9858061-A1.
PD   23-DEC-1998.
PF   18-JUN-1998; U12236.
PR   09-OCT-1997; US-061641.
PR   19-JUN-1997; US-878730.
PR   19-JUN-1997; US-878878.
PA   (SCHE ) SCHERING CORP.
PI   Franz-Bacon K, Gorman DM, McClanahan TK;
DR   WPI; 99-095339/08.
DR   P-PSDB; W87709.
PT   New cysteine-rich soluble proteins - used to modulate proliferation,
PT   differentiation, trafficking and development of cells, e.g. for
PT   treating inflammation, cancer and degeneration
PS   Claim 16; Page 17; 119pp; English.
CC   The present sequence encodes a cysteine rich soluble protein (CRSP)
CC   designated C10. CRSP proteins, and their (ant)agonists, are used to
CC   modulate physiology, differentiation, trafficking and development
CC   of cells (including those in culture), particularly cells of the
CC   mammalian immune system. They are used for treatment of abnormal
CC   proliferation (cancer, inflammation or degeneration), regeneration,
CC   degeneration and atrophy. The proteins are also used to raise, or
CC   detect, antibodies, to design oligonucleotides for library screening,
CC   in drug screens and to isolate cognate receptors. The antibodies are
CC   used for affinity purification of CRSP, to screen expression libraries,
CC   to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC   to produce anti-idiotypic antibodies (useful for diagnosis), in
CC   competitive drug screens, and as therapeutic modulators.
SQ   Sequence 603 BP;    149 A;    170 C;    135 G;    149 T;

Query Match        20.5%;   Score 92.8;   DB 1;   Length 603;
Best Local Similarity  58.3%;   Pred. No. 4.1e-15;
Matches 182;   Conservative 0;   Mismatches 127;   Indels 3;   Gaps 1;

Qy    56 CTCTGTCTCCTCCTCCCTCCCTGTCCTCGGGGCTGTTGGTGTGTTAGCAAGACCCTGTGCTCC 115
          |||  |||  ||| ||| | |||||| |||| ||| | ||||| |||| || | | ||||||
Db   126 CTCCTTCTCATCCTAAATCCCCCTTCCCAGCTGGGCTCCAGCTGATCAACCCGGGAGTGTGTTCC 185

Qy   116 ATGGAAGAAGCCATCAATGAGAGGATCCAGGAGGTCGCCGGCTCCCAATATTAGG--- 173
          |||  || |||||| ||||  ||||||||| | | ||||| |||||| |||||
Db   186 TTAGACTCCGTTATGATAAGAAGATCAAGAGATGTTCTCAACAGTCTAGAGTACAGTCCC 245

Qy   173 GCAATAAGCAGCATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTACT 232
          ||||| ||  |||||||| |  || ||| | ||| |  |  ||| ||| || | |
Db   246 TCTCCCATAAGCAAGAAGCTCTCGTTCGTTGCTAGTTGTCAAAAGCCAAGGCAGACCGTCCTCC 305

Qy   233 TGCCCCCGAGGCTTCGCCCGTCACCGGCTGCACTTGGGCTGTGTTGCTCAAGACCCTGTGTCC 292
          ||||||  | | | || | || || || ||| | | |||||| | |||| | ||| ||||
Db   306 TGCCCCTGCTGGGATGGCGTGCTGCTCACTGGCTGTGTGCTATGGCTGTGGTTCGTGGGAT 365

Qy   293 GTGCGGCCGGAGACCACCATGCTCCAGTGCGCGGGCATGGACTGGACGGACGCGCGC 352
          ||  | ||| || | ||||| | |||| |||  |  ||||| |||| |||  | | ||
Db   366 GTTCAGCTGGAAACACCTGCCACTGCCACTGCCAGTGTGGGACTGGACCACTGCCCCGC 425

Qy   353 TGCTGTCGTGTG 364
          ||||| | | ||
Db   426 TGCTGCCACCTG 437
```

```
FT                   /note= "intron present between these nucleotides in
FT                          the genomic DNA"
PN   WO9858061-A1.
PD   23-DEC-1998.
PF   18-JUN-1998; U12236.
PR   09-OCT-1997; US-061641.
PR   19-JUN-1997; US-878730.
PR   19-JUN-1997; US-878878.
PA   (SCHE ) SCHERING CORP.
PI   Franz-Bacon K, Gorman DM, McClanahan TK;
DR   WPI; 99-095339/08.
DR   P-PSDB; W87707.
PT   New cysteine-rich soluble proteins - used to modulate proliferation,
PT   differentiation, trafficking and development of cells, e.g. for
PT   treating inflammation, cancer and degeneration
PS   Claim 16; Page 15; 119pp; English.
CC   The present sequence encodes a cysteine rich soluble protein (CRSP)
CC   designated C19. CRSP proteins, and their (ant)agonists, are used to
CC   modulate physiology, differentiation, trafficking and development
CC   of cells (including those in culture), particularly cells of the
CC   mammalian immune system. They are used for treatment of abnormal
CC   proliferation (cancer, inflammation or degeneration), regeneration,
CC   degeneration and atrophy. The proteins are also used to raise, or
CC   detect, antibodies, to design oligonucleotides for library screening,
CC   in drug screens and to isolate cognate receptors. The antibodies are
CC   used for affinity purification of CRSP, to screen expression libraries,
CC   to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC   to produce anti-idiotypic antibodies (useful for diagnosis), in
CC   competitive drug screens, and as therapeutic modulators.
SQ   Sequence 560 BP;    137 A;    160 C;    132 G;    131 T;

Query Match        33.6%;   Score 152.4;   DB 1;   Length 560;
Best Local Similarity  67.8%;   Pred. No. 3.1e-30;
Matches 213;   Conservative 0;   Mismatches 101;   Indels 0;   Gaps 0;

Qy    54 CTCTGTCTCCTCCTCCCTCCCTGTCCTCGGGGCTGTTGGTGTGTTAGCAAGACCCTGTGCT 113
          | ||| |  || |||| ||||||||||| ||| | ||||| ||| | || | | |||||
Db    83 CCCTCCTTTCTTTCTTCCTTGTCCCTGAACTGCTGGGCTCCAGCATGCCACTGTGTC 142

Qy   114 CCATGGAAGAAGCCATCAATGAGAGGATCCAGGAGGTCGCCGGCTCCCAATATTAGG 173
          |||||||| ||||||||||||||||||| | | ||||| |  ||||||||||
Db   143 CCATCGATGAAGCCATCGACAAGAAGATCAACAAGAGACTTCAACTCCCTGTTTCCAAATG 202

Qy   174 CAATAAGCAGCATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTACTT 233
          |||||| |  ||||||||| |  ||||| ||||| |  || ||| ||| ||| | |
Db   203 CAATAAAGACAACATTGGCTTAAATTGCTGGACAGTCTCCCCAGAGGGAGTTGGCCTCCT 262

Qy   234 GCCCCCGAGGCTTCGCCGTCACCGGCTGCACTTGGCCTCCGCCTGCTGCGCCTTGCGGGATG 293
          ||||  | | | || | || || || ||| | |
Db   263 GCCCAGAGGCCACAGCAGTCTTGACGTCTGCTCCTGGCTCGTGCGCGGCTCGTGGGACA 322

Qy   294 TGCGGCCGAGACCACCATGCTCCAGTGCGCGGGCATGGACTGGACGGACGCGCGCT 353
Db   323 TTCGTGAAGAAAAGTGTGTCACTGCCAGTGCCAAGGATAGACTGGACAGCAGCCCGCT 382

Qy   354 GCTGTCGTGTGCAG 367
          |||||    | |||
Db   383 GCTGTAAGCTGCAG 396
```

```
RESULT 6
V84058
ID   V84058 standard; cDNA; 603 BP.
AC   V84058;
DT   09-MAR-1999 (first entry)
DE   cDNA encoding a cysteine rich soluble protein; CRSP; C10; cell development.
KW   Cysteine rich soluble protein; antibody;
KW   mammalian immune system; antibody; abnormal proliferation; cancer;
KW   inflammation; degeneration; regeneration; degeneration; atrophy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             108..443
```

RESULT 7
V84055
ID   V84055 standard; cDNA; 554 BP.
AC   V84055;
DT   09-MAR-1999 (first entry)
DE   cDNA encoding a cysteine rich soluble protein designated C18.
KW   Cysteine rich soluble protein; CRSP; C18; cell development;
KW   mammalian immune system; antibody; abnormal proliferation; cancer;
KW   inflammation; degeneration; regeneration; atrophy; ss.
OS   Mus sp.
FH   Key   Location/Qualifiers
FT   CDS   103..420
FT   /tag= a
FT   /product= C18
FT   sig_peptide   103..159
FT   /tag= b
FT   mat_peptide   160..417
FT   /tag= c
FT   misc_feature   230..231
FT   /tag= d
FT   /note= "intron present between these nucleotides in
FT   genomic DNA"
FT   misc_feature   292..293
FT   /tag= e
FT   /note= "intron present between these nucleotides in
FT   genomic DNA"
PN   WO9858061-A1.
PD   23-DEC-1998.
PF   18-JUN-1998; U12236.
PR   09-OCT-1997; US-061641.
PR   19-JUN-1997; US-878730.
PR   19-JUN-1997; US-878878.
PA   (SCHE) SCHERING CORP.
PI   Franz-Bacon K, Gorman DM, McClanahan TK;
WPI: 99-095339/08.
-DR P-PSDB; W87706.
PT   New cysteine-rich soluble proteins - used to modulate proliferation,
PT   differentiation, trafficking and development of cells, e.g. for
PT   treating inflammation, cancer and degeneration
PS   Claim 16; Page 14; 119pp; English.
CC   The present sequence encodes a cysteine rich soluble protein (CRSP)
CC   designated C18. CRSP proteins, and their (antagonists), are used to
CC   modulate physiology, differentiation, trafficking and development
CC   of cells (including those in culture), particularly cells of the
CC   mammalian immune system. They are used for treatment of abnormal
CC   proliferation (cancer, inflammation or degeneration), regeneration,
CC   degeneration and atrophy. The proteins are also used to raise, or
CC   detect, antibodies, to design oligonucleotides for library screening,
CC   in drug screens and to isolate cognate receptors. The antibodies are
CC   used for affinity purification of CRSP, to screen expression libraries,
CC   to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC   to produce anti-idiotypic antibodies (useful for diagnosis), in
CC   competitive drug screens, and as therapeutic modulators.
SQ   Sequence 554 BP; 131 A; 138 C; 130 G; 155 T;

Query Match   17.2%; Score 78; DB 1; Length 554;
Best Local Similarity   56.5%; Pred. No. 2.3e-11;
Matches   175; Conservative   0; Mismatches 120; Indels   15; Gaps 1;

Qy   55 TCTCGTCTCCTCCTCCTCCTGTCCTCGGGCGTTGGTGTCTAGCCAAGACCCTGGCTC 114
Db 120 TTTCCTTTTCATCCTCGTCCCCTTTCCCACTGATAGTCCCAGGGAACGCGCAATGCTC 179
Qy 115 CATGGAAGAAGCCATCAATGAGAGGATCCAGGAGGTCGCCGGCTCCCTAATATTTAGGGC 174
Db 180 CTTTGAGTCTTGTGGATCAAAGGATCAAGGAAGCTCTCAGTCGTCAAGAGCCTAAGAC 239
Qy 175 AATAAGCAGCATTGGCCTGAGTGCCAGAGCGCTCACCTCCAGGGGGACCTGGCTACTTG 234
Db 240 GATCTCCTGCACT---------AGTGTCACGTCTTCTGGCAGACTGGCCTCCTG 284
Qy 235 CCCCCGAGGCTTCGCGCTCACCGGCTCACCTTGTGGCCTCCCCGTTGTCCCGTCCCTGT 294

Db 285 TCCTGCTGGATGTTGTCACTGGATGTGCTGTTGTGCTATGCTGTGGATCGTGGGATAT 344
Qy 295 GCGCGCGAGACCACCATGTCACTCCCAGTGCGCGGGCCATGGACGACTGGACGCCGGCTG 354
Db 345 CCGGAATGGAAAATACTTGCCACTGCCAGTGCTCAGTCATGGACTGGCCTTCTGCCCGCTG 404
Qy 355 CTGTCGTGTG 364
Db 405 CTGCCGAATG 414

RESULT 8
V84054
ID   V84054 standard; cDNA; 574 BP.
AC   V84054;
DT   09-MAR-1999 (first entry)
DE   cDNA encoding a cysteine rich soluble protein designated C2b.
KW   Cysteine rich soluble protein; CRSP; C2b; cell development;
KW   mammalian immune system; antibody; abnormal proliferation; cancer;
KW   inflammation; degeneration; regeneration; atrophy; ss.
OS   Mus sp.
FH   Key   Location/Qualifiers
FT   CDS   70..405
FT   /tag= a
FT   /product= C2b
FT   sig_peptide   70..138
FT   /tag= b
FT   mat_peptide   139..402
FT   /tag= c
FT   misc_feature   196..197
FT   /tag= d
FT   /note= "intron present in genomic DNA"
FT   misc_feature   277..278
FT   /tag= e
FT   /note= "intron present in genomic DNA"
PN   WO9858061-A1.
PD   23-DEC-1998.
PF   18-JUN-1998; U12236.
PR   09-OCT-1997; US-061641.
PR   19-JUN-1997; US-878730.
PR   19-JUN-1997; US-878878.
PA   (SCHE) SCHERING CORP.
PI   Franz-Bacon K, Gorman DM, McClanahan TK;
WPI: 99-095339/08.
DR   P-PSDB; W87705.
PT   New cysteine-rich soluble proteins - used to modulate proliferation,
PT   differentiation, trafficking and development of cells, e.g. for
PT   treating inflammation, cancer and degeneration
PS   Claim 16; Page 13; 119pp; English.
CC   The present sequence encodes a cysteine rich soluble protein (CRSP)
CC   designated C2b. CRSP proteins, and their (ant)agonists, are used to
CC   modulate physiology, differentiation, trafficking and development
CC   of cells (including those in culture), particularly cells of the
CC   mammalian immune system. They are used for treatment of abnormal
CC   proliferation (cancer, inflammation or degeneration), regeneration,
CC   degeneration and atrophy. The proteins are also used to raise, or
CC   detect, antibodies, to design oligonucleotides for library screening,
CC   in drug screens and to isolate cognate receptors. The antibodies are
CC   used for affinity purification of CRSP, to screen expression libraries,
CC   to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC   to produce anti-idiotypic antibodies (useful for diagnosis), in
CC   competitive drug screens, and as therapeutic modulators.
SQ   Sequence 574 BP; 161 A; 139 C; 122 G; 152 T;

Query Match   14.2%; Score 64.4; DB 1; Length 574;
Best Local Similarity   62.3%; Pred. No. 6.5e-08;
Matches   101; Conservative   0; Mismatches   61; Indels   0; Gaps 0;

Qy 203 AGCGTCACCTCCAGGGGGACCTGGCTACTTGCCCCCGAGGCTTCGCCCCAGAGCGTCCGCTCCACCGGCTGC 262
Db 238 AGTATCACGGCTTCAGGCAGACTGGCGCTCCGTCCTTCGGAATGACTGCTGTCACTGGTTGT 297

```
Qy    263 ACTTGTGGCTCCGCCTGGCTCGTGGGATGTGCGCCGCCGAGACCACATGTCACTGCCAG 322
          |||||||| |  ||||||| | ||||||  | |||||| |  | | |||||| |||| |||
Db    298 GCTTGTGGCTATGGCTGTGGATCTTGGGATATCCGGGGATGGAAATACTTGCCACTGTCAG 357

Qy    323 TGCGCGGGCATGGACTGACCGGAGCGCGCTGCGTGTCGTGTG 364
          |||  ||||||||| || |||||||| || ||  ||| ||
Db    358 TGCTCAACAATGGACTGGGCCACCGCCCGTTGCCACTGCCAACTG 399


RESULT    9
V84053
ID   V84053 standard; cDNA; 527 BP.
AC   V84053;
DT   09-MAR-1999 (first entry)
DE   cDNA encoding a cysteine rich soluble protein designated C2.
KW   Cysteine rich soluble protein; CRSP; C2; cell development;
KW   mammalian immune system; antibody; abnormal proliferation; cancer;
KW   inflammation; degeneration; regeneration; degeneration; atrophy; ss.
OS   Mus sp.
FH   Key            Location/Qualifiers
FT   CDS            32..367
FT                  /*tag= a
FT                  /product= C2
FT   sig_peptide    32..100
FT                  /*tag= b
FT   mat_peptide    101..364
FT                  /*tag= c
PN   WO9858061-A1.
PD   23-DEC-1998.
PF   18-JUN-1998; U12236.
PR   09-OCT-1997; US-061641.
PR   19-JUN-1997; US-878730.
PR   19-JUN-1997; US-878878.
PA   (SCHE ) SCHERING CORP.
PI   Franz-Bacon K, Gorman DM, McClanahan TK;
DR   WPI; 99-095339/08.
DR   P-PSDB; W87704.
PT   New cysteine-rich soluble proteins - used to modulate proliferation,
PT   differentiation, trafficking and development of cells, e.g. for
PT   treating inflammation, cancer and degeneration
PS   Claim 16; Page 12-13; 119pp; English.
CC   The present sequence encodes a cysteine rich soluble protein (CRSP)
CC   designated C2. CRSP proteins, and their (ant)agonists, are used to
CC   modulate physiology, differentiation, trafficking and development
CC   of cells (including those in culture), particularly cells of the
CC   mammalian immune system. They are used for treatment of abnormal
CC   proliferation (cancer, inflammation or degeneration), regeneration,
CC   degeneration and atrophy. The proteins are also used to raise, or
CC   detect, antibodies, to design oligonucleotides for library screening,
CC   in drug screens and to isolate cognate receptors. The antibodies are
CC   used for affinity purification of CRSP, to screen expression libraries,
CC   to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC   to produce anti-idiotypic antibodies (useful for diagnosis), in
CC   competitive drug screens, and as therapeutic modulators.
SQ   Sequence 527 BP; 135 A; 126 C; 116 G; 150 T;

Query Match    12.0%;      Score 54.4;   DB 1;   Length 527;
Best Local Similarity 57.7%;      Pred. No. 2.2e-05;
Matches 97; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy    197 TGCCAGAGCGTCACCTCCAGGGGGACCTGCTACTTGCCCCCGAGGCTTCGCCGTCACC 256
          ||| |||| | | ||||   |||  | || || | ||| ||||||| | ||||
Db    194 TGCACTAGTGTCAAGACTATGAACAGATGGGCCCTCCTGCCCTGGGATGACTGCTACT 253

Qy    257 GGCTGCACTTGTGCCTCCGCCTGTGGCCTGTGGATGTGCGCGCCGAGACCACCATGTCAC 316
          |||||  | || | | |  ||| |||| |||| ||| | || |||  |  ||||| |
Db    254 GGGTGCGTCGTTGTGCCTTGCCTTGGATCTTGGGAGATCTTGGGAGATGGAGATACTTGCAAC 313

Qy    317 TGCCAGTGCGGCCGGCATGGACTGGACCGGACCGCCCGGCTGCTCGTCGTG 364
          ||| ||| |  | || |  | | ||||||  |||||| | |||| ||  |
Db    314 TGCCTGTGCTACTCGTTGACTGGACCACTGCCGCCCGGCTGCCCAACTG 361


RESULT    11
V64548
ID   V64548 standard; DNA; 985 BP.
```

```
RESULT    10
V44439
ID   V44439 standard; DNA; 985 BP.
AC   V44439;
DT   09-NOV-1998 (first entry)
DE   Mycobacterium tuberculosis antigen XP25 3' DNA.
KW   Tuberculosis; infection; diagnosis; antigen; XP25; ss.
OS   Mycobacter tuberculosis strain Erdman.
PN   WO9816645-A2.
PD   23-APR-1998.
PF   07-OCT-1997; U18214.
PR   13-MAR-1997; US-818111.
PR   11-OCT-1996; US-729622.
PA   (CORI-) CORIXA CORP.
PI   Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI   Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
DR   WPI; 98-251292/22.
PT   New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT   to develop products for the detection of M. tuberculosis infection
PT   and diagnosis of tuberculosis
PS   Claim 4; Page 187-188; 250pp; English.
CC   This is the 3' region of DNA coding for an antigenic portion of
CC   Mycobacterium tuberculosis antigen XP25; 5' DNA is provided in
CC   V4443B. XP25 DNA was isolated from a M. tuberculosis strain Erdman
CC   genomic DNA expression library using sera from patients having
CC   extrapulmonary tuberculosis. It bears no similarity to known
CC   sequences. The invention relates to methods and compositions for
CC   diagnosing tuberculosis. It provides polypeptides (see
CC   W64291-W64379) comprising an antigenic portion of a soluble M.
CC   tuberculosis antigen, or an immunogenic portion of a M.
CC   tuberculosis antigen, as well as DNA sequences encoding such
CC   polypeptides, recombinant expression vectors and transformed or
CC   transfected host cells. Also claimed are methods and diagnostic
CC   kits for detecting M. tuberculosis infection in a patient using
CC   these polypeptides, antibodies or oligonucleotide probes and
CC   primers.
SQ   Sequence 985 BP; 138 A; 330 C; 423 G; 94 T;

Query Match    9.0%;      Score 40.6;   DB 1;   Length 985;
Best Local Similarity 45.3%;      Pred. No. 0.079;
Matches 148; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

Qy    111 GCTCCATGGAAGAAGCCATCAATGAGGATCCAGGAGGTCGCCGGCTCCCTAATATTTA 170
          | || ||| |  |  | |  |  | || |  ||| || ||||| || || ||||| ||
Db     98 GCGCCCGGCGGTGCAGGCGGCCAAGGGCGGCCACCGGAGGTCGCCGCGGGCTCCGGCGGGACA 157

Qy    171 GGGCAATAAGCAGCATTGGCCTGGAGTGCCGAGAGTCACCTCCAGGGGGGACCTGGCTA 230
          ||| ||| | | || |  || ||  | ||| | | || || | ||||||  || || |
Db    158 ACCCCACCGGTGCTCGTTTCGCCGGTGGCCGGTGGGCGGCGCACCAGGTGGCGCGGCCGCCGG 217

Qy    231 CTTGCCCCGAGGCTTCGCCGTCACCTGCACTTGTGGCGCGGCCGCTGGCGTCGTGGG 290
          | || | |||||| || |||| | || ||| || || | |  | | | |||| |
Db    218 GCGGGGCCGCGGGGACGGGGGTACCGGCGGCCGGGGGCCCGGGCATGGACCGGCGGACCGCGC 277

Qy    291 ATGTGCGCGCCGAGACCACCACATGTCACTGCCAGTGCGGCATGGACTGGCGGACGGCCGC 350
          |||| | |  || |  | |  | ||| | ||| ||| | || ||||| |||| | |||| |
Db    278 GTGCAGGCATCGGCGGCCCGCGGGGTGACGGCCGGCGGCGGCGGGGCAGCGGCGTCAGCGGTC 337

Qy    351 GCTGCTGCTGTGTGTGAGCTCCGGGGACCTGGCCGTCGAGGTCGCCCGAGGTCGCCCTGCCACC 410
          ||| | ||| | | |  | ||| ||| || || | |||   ||| ||  |  || ||||
Db    338 TCGGCCTGGCCGCCGTTGACGGCGGCGGCCGGCCAAGGCGGCGGCCAAGGGCGGCGGGCGCGGCA 397

Qy    411 GCTCCAGGTCCGAGGGGGTTGCGGGGG 437
          ||| | |  | ||||| ||  | ||
Db    398 GCGGCGGCCGCGCGGCCATCAACGGGG 424
```

AC   V64548;
DT   27-JAN-1999 (first entry)
DE   M. tuberculosis immunogenic polypeptide XP25 3'-end DNA.
KW   Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW   vaccine; pharmaceutical; infection; diagnosis; ss.
OS   Mycobacterium tuberculosis.
PN   WO9816646-A2.
PD   23-APR-1998.
PF   07-OCT-1997; U18293.
PR   13-MAR-1997; US-818112.
PR   11-OCT-1996; US-730510.
PA   (CORI-) CORIXA CORP.
PI   Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI   Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
DR   WPI; 98-261042/23.
PT   Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT   to develop products for the detection of M. tuberculosis infection
PT   and for diagnosis, treatment and prevention of tuberculosis
PS   Claim 4; Page 176; 230pp; English.
CC   This sequence encodes an immunogenic portion of a soluble Mycobacterium
CC   tuberculosis (MT) antigen which can be used in a method for inducing
CC   protective immunity against tuberculosis (TB). This sequence can be
CC   formulated into vaccines and/or pharmaceutical compositions for
CC   immunising against M. tuberculosis infection or may be used for the
CC   diagnosis of tuberculosis.
SQ   Sequence 985 BP; 138 A; 330 C; 423 G; 94 T;

Query Match   9.0%; Score 40.6; DB 1; Length 985;
Best Local Similarity 45.3%; Pred. No. 0.079;
Matches 148; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

Qy 111 GCTCCATGGAGAAGCCATCAATGAGAGGATCCAGGAGGTCGCCGGCTCCCTAATATTA 170
         || || ||| ||| | | | |||| | || |||| ||
Db   98 GCGCCGGCCGGTGCAGGCGCCGCAAGGCGCGGTGCCGGGTCCGGCGCGGGACA 157

Qy 171 GGGCAATAAGCAGCATTGGCCTGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTA 230
         | | | |||| ||| | | | | ||| || | | |||| || |
Db 158 ACCCCACCCGGTGCTGGTTTCGCCGGTGGCGGGCGGGCGCACAGGTGGCCGCCGGCGCG 217

Qy 231 CTTGCCCCGAGGCTTCGCCCGTCACCGGTGCACTTGTGGTCCTCCGCCTGTGGCTCGTGGG 290
         ||| || | | ||||| ||| ||| |||
Db 218 GCGGGGCCGGCCGGCGACCGGTACCGGCGCACCCGGCGCGTTGTCGGCGACTGGACGGCGC 277

Qy 291 ATGTGCGCGCGAGACCACATGTCACTGCCAGTGCCCGGCGGGGCCGGCGGGGGCATGGGGC 350
         | | | | | | | || ||| ||| || || |
Db 278 GTGCAGGCATCGGGCGCCGGCCGAGGCCCCGAGGGCCGGGTGACGGGCGGGCCAGCGGTC 337

Qy 351 GCTGCTGTGTGTGCAGCCCTGAGTGCCGCGCGCAGCCGGTGCACAGCGGCGGGCGGAGGCG 410
         || || | | ||| | | ||| |||| || || |
Db 338 TCGGCCTGGGCCCTCCGGCTTTGACGGCGCGGCCAAGGCGGGCCGGCCAAGGCGGGCGGCA 397

Qy 411 GCTCCAGGTCCGGAGGGGGTTGCGGGGG 437
         || || || |||| || |
Db 398 GCGCCGGCCGGCGGCATCAACGGGG 424

RESULT 12
Q73500
ID   Q73500 standard; DNA; 8438 BP.
AC   Q73500;
DT   15-MAY-1995 (first entry)
DE   DNA encoding Pseudorabies virus large latency transcript.
KW   Pseudorabies virus; PRV; LLV; large latency transcript;
KW   attenuated virus; vaccine; early protein 0; EP0; HSV-1 ICP0;
KW   protecting animals; deletion mutants; swine; ds.
OS   Pseudorabies virus.
FH   Key   Location/Qualifiers
FT   misc_feature   1..7013
FT   /tag= a
FT   /note= "derived from PRV strain InFh"
FT   misc_feature   7014..8425
FT   /tag= b

FT   cds   .
FT   /note= "derived from PRV strain Ka"
FT   622..6498
FT   /*tag= c
FT   /note= "encodes predicted amino acid sequence of ORF2"
FT   tata_signal   1..6
FT   /*tag= d
FT   misc_feature   34
FT   /*tag= e
FT   /note= "RNA cap site"
FT   polya_signal   8382..8387
FT   /*tag= f
PN   US5352596-A.
PD   04-OCT-1994.
PF   11-SEP-1992; 945283.
PR   11-SEP-1992; US-945283.
PA   (USDA) US SEC OF AGRIC.
PI   Cheung AK, Wesley RD;
DR   WPI; 94-316187/39.
DR   P-PSDB; R60620.
PT   New pseudorabies virus mutants for use in vaccine - having a
PT   deletion and/or insertion in the early protein O gene or large
PT   latency transcript gene
PS   Disclosure; Column 15-30; 43pp; English.
CC   Q73500 shows the Pseudorabies virus (PRV) large latency transcript
CC   (LLT). The basic sequence is derived from PRV strain InFh and PRV
CC   strain Ka. The LLT overlaps and is transcribed in the opposite
CC   orientation with respect to the EP0 (early polypeptide 0) and the
CC   immediately early gene (IE180). EP0 is nonessential for replicatio,
CC   LLT is the only gene expressed during PRV latency, and the IE180
CC   gene is absolutely necessary for PRV replication. However there are
CC   2 copies of IE180 in the genome. It is expected that PRV lacking one
CC   of the IE180 copies is viable. Deletions in the non-overlapping
CC   regions of these 3 genes will generate single deletion routants,
CC   while deletions in overlapping regions will generate double deletion
CC   mutants. The invention is concerned with the construction of attenuated
CC   viruses which have a reduced ability to reactivate from latency. This
CC   can be achieved by functionally disabling the expresion of the EP0
CC   gene, or by disrupting the synthesis of the LLT, or both.(See also
CC   Q73501 and R60620-24)
SQ   Sequence 8438 BP; 1141 A; 2916 C; 3327 G; 1054 T;

Query Match   8.6%; Score 39; DB 1; Length 8438;
Best Local Similarity 51.6%; Pred. No. 0.31;
Matches 114; Conservative 0; Mismatches 105; Indels 2; Gaps 1;

Qy 217 GGGGGACCTGGCTACTTGCCCCGAGGCTTCGCCCGTCACCGGTGCCCTCACTTGTGGTCCGC 276
         || ||| | | || || |||||||||| || | | || ||||| ||||||||||
Db 5479 GGCGGAGGAGGCCCGAGGGCCCGGCGGGCCCGGCCGGGAGACCGTGGCGGCCCC 5538

Qy 277 CTGTGGCTCGTGGGATGTGCGCGCGAGACCACATGTCACTGCCAGTGCGCGGCATGGA 336
         ||| | | | | || | | | | || || ||
Db 5539 GGGCGGCGGCGAGTGGGGCGCCGGCCGGACTCCTTGCGTCTTCGTCTTCTCCCTCCGGAGGAGGA 5598

Qy 337 CTGGACC--GGAGCGCGCGTGCTGTGTGCAGCCCTGAGGTCCGTGTGCAC 394
         ||||| ||| | | | | |
Db 5599 CGAGGACGAGGAGGAGGACGAGGAGGACGAGGAGGACCGGCGGCCGAGCGCCGCGCGGC 5658

Qy 395 AGCGGCGGGCGGAGGGCGGGCCGGCTCCAGGTCCGGAGGGGTTGCGGG 435
         ||| || | || || |||| ||| | || || ||
Db 5659 GGCGGCGGCGGGCCGGCGGGGGCCGGGGGCCGAGGGGGCGGAGGCGCGCGCGGC 5699

RESULT 13
X53491
ID   X53491 standard; DNA; 114955 BP.
AC   X53491;
DT   05-JUL-1999 (first entry)
DE   Human adenosine A1 receptor antisense oligonucleotide fragment.
KW   Antisense oligonucleotide; multiple target; antisense treatment;
KW   impaired respiration; inflammation; lung disease;
KW   pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW   acute asthma; allergy; asthma; impeded respiration;

```
KW   respiratory distress syndrome; pain; cystic fibrosis;              KW   immunoassay; detection; mRNA; assay; ss.
KW   pulmonary hypertension; pulmonary vasoconstriction; emphysema;      OS   Homo sapiens.
KW   chronic obstructive pulmonary disease; leukemia; lymphoma;          FH   Key                Location/Qualifiers
KW   colon cancer; breast cancer; lung cancer; pancreatic cancer;        FT   cds                337..2736
KW   hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;  FT                      /*tag= a
KW   prostate cancer; ss.                                                FT   signal_peptide     337..405
OS   Synthetic.                                                          FT                      /*tag= b
PN   WO9913886-A1.                                                       FT   mat_peptide        406..2733
PD   25-MAR-1999.                                                        FT                      /*tag= c
PF   17-SEP-1998; U19419.                                                PN   US5527679-A.
PF   09-JUN-1998; US-093972.                                             PD   18-JUN-1996.
PR   17-SEP-1997; US-059160.                                             PF   01-MAY-1991; 694314.
PA   (UYEC-) UNIV EAST CAROLINA.                                         PR   01-MAY-1991; US-694314.
PI   Nyce JW;                                                            PR   27-APR-1993; US-054077.
DR   WPI; 99-229400/19.                                                  PA   (DAND ) DANA FARBER CANCER INST INC.
PT   New antisense oligonucleotides used in treatment of, e.g. pulmonary  PI   Hemler ME; Ramaswamy H;
PT   vasoconstriction                                                    DR   WPI; 96-299852/30.
PS   Disclosure; Page 37; 120pp; English.                               DR   P-PSDB; W02194.
CC   The specification describes antisense oligonucleotides (X52869-X55271)  PT   Immunoassay and mRNA hybridisation assay for beta-5 protein - useful
CC   directed against at least 2 mRNAs selected from target genes, coding and  PT   for the detection of carcinoma(s) and to distinguish different cell
CC   non-coding regions of RNAs corresponding to target genes, gene     PT   types
CC   initiation codons, genomic flanking regions, intron-exon borders, the  PS   Claim 11; Columns 13-20; 21pp; English.
CC   5'-end, the 3'-end and the juxta-section between coding and non-coding  CC   The present sequence encodes the human integrin beta subunit
CC   regions and all segments of RNAs encoding proteins associated with one  CC   protein, beta-5, which is found in carcinomas but not in lymphoid
CC   or more diseases, conditions or mixtures. The antisense oligonucleotides  CC   cells. An immunoassay for the detection of beta-5, comprises
CC   may be derived from sequences X55272-74. These multiple target     CC   contacting a sample with a monoclonal antibody (Ab) which binds 1
CC   oligonucleotides (specifically X55180-271) can be used for the antisense  CC   epitope of beta-5, and then with a labelled Ab which binds another
CC   treatment of diseases and conditions. Typical diseases and conditions  CC   epitope of beta-5, and detecting any bound label. An assay for
CC   are those associated with impaired respiration and inflammation,   CC   beta-5 mRNA, comprises contacting a sample with a probe capable of
CC   including lung diseases, pulmonary vasoconstriction, inflammation,  CC   hybridising to the beta-5 cDNA, and determining if binding has
CC   allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,  CC   occurred. These assays are useful for detecting carcinomas, and for
CC   respiratory distress syndrome, pain, cystic fibrosis, pulmonary    CC   distinguishing between different cell types.
CC   hypertension, pulmonary vasoconstriction, chronic obstructive      SQ   Sequence  3415 BP;  752 A;  943 C;  1009 G;  711 T;
CC   pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC   carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic  Query Match             8.1%;  Score 36.6;  DB 1;  Length 3415;
CC   cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic  Best Local Similarity  47.6%;  Pred. No.1;
CC   metastases, as well as all types of cancers which may metastasize or have  Matches  108;  Conservative   0;  Mismatches  119;  Indels   0;  Gaps    0;
CC   metastasized to the lungs, including breast and prostate cancer.
SQ   Sequence  114955 BP;  6071 A;  29417 C;  36712 G;  21328 T;          Qy   140 ATCCAGGAGGTCGCCGGCTCCCTAATATTAGGCAATAAGCAGCATTGGCCTGGAGTGC 199
                                                                            |||||||| ||||||||||| |  ||||  | ||  | |||| |||||| || ||||
Query Match             8.5%;  Score 38.6;  DB 1;  Length 114955;       Db  1690 AGCCTGGAGTGGGGTCACTTCCAGGGGGACGTGGCTACTTGCCCCCGAGGCTTCGCCGCTGCCCGGGTGGGCGTCACCGGC 1749

Best Local Similarity  36.0%;  Pred. No. 0.65;                         Qy   200 CAGAGCGTCACCTCCAGGGGACGTGGCTACTTGCCCCC 239
Matches  93;  Conservative   19;  Mismatches  146;  Indels   0;  Gaps    0;   |||||| ||||||  ||| |  |||| |||||| |
                                                                       Db  1750 AACAGCGCCAGGTGCAACGGGAGCGGGACCTATGTCGCGGGCCGCTGTGGGATGTCGCGCGCCGAGACCCACCATGTCACTGC 1809
Qy   180 GCAGCATTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTACTTGCCCCC 239
          ||||||||||      |||||    ||||||  |||||                     Qy   260 TGCACTTGTGGCTCCGCTCGTGGGCTCCGCCTGTGGCTCGTGGGATGTGCGGG 319
Db 103818 GCGGGCGTCGGGGGCCGGGGCCGGGGCSNNNDNNCGTC 103877                        |||| |  | | ||   |||||| | | | |||  |||| |  |||||| ||
                                                                       Db  1810 GGCTACCTGGGCACCAGGTGCCAGAGTGCCAGGATGGGAGAACCAGGAGCGTGTGTACCAGAAC 1869
Qy   240 GAGGCTTCGCCGCTCACCGGCTCCGGCCGGGCSNNNDNNGTGGGCTCCGCCTGTGGCTCGTGGGATGTGCGGG 299
          ||   ||   | ||     |||     | ||  | ||| |||                  Qy   320 CAGTGCGCGGCATGGACTGACCGGACCGGGCGCTGCTGTGTCGTGCA 366
Db 103878 BTGGCGGCGGCGTCGGCCGGCCGGGCCGGGCSNNNDNNGTCBTGGCGGCGGCSNNNDNNTCB 103937           :||  |||  ||  |||||| |||| | | || ||  ||||
                                                                       Db  1870 CTGTGCCGGGAGGCAGCAGAGGGCAAGCCCACTGTGCCAGCCGGCCGGCCGTGGGGA 1916
Qy   300 CCGAGACCACCATGTCACTGCCAGTGCGGGGCCATGGCGGGGCCGGGGCCGGGGCSNNNDNNBTGGCG 359
          |||     ||| |||||  || |   ||:    ||||||  |:     |    | |||| |||||
Db 103938 TGGCGGCGTCGGCCGGCCGGGCSNNNDNNCBTGGCGGCGGCCGGGCSNNNDNNBTGGCG 103997  RESULT 15
                                                                       V18130/c
Qy   360 GTGTGCAGCCCTGAGGTCGGCGGCCGCAGCCGGTGCACAGCCGGGCGGAGGCGGTTCCAGGT 419  ID   V18130 standard; DNA; 11820 BP.
          | | |||   ||| ||| ||   | ||  :   | ||    |||  |::| |           AC   V18130;
Db 103998 GGCTCGGGGGCCGGGGCCGGGGCSNNNDNNTGGCGGGCCGGGCCTCGGGGCCGCGGCCGTCGGGC 104057  DT   04-SEP-1998 (first entry)
                                                                       DE   Human chromosome 19 derived USF2 gene sequence.
Qy   420 CCGGAGGGGTGCGGGGG 437                                          KW   Hydronephrosis gene; HNG gene; USF2 gene; renal disease; renal aplasia;
          ||::  |||  |||||                                            KW   vesical-ureteral reflux; pelvi-ureteral junction obstruction;
Db 104058 CGGGCSNNNDNNGCGGCG 104075                                     KW   multicystic renal dysplasia; renal agenesis; hydronephrosis;
                                                                       KW   Von Mayer-Rokitansky-Kuester disorder; bifid ureter; ss.
RESULT 14                                                              OS   Homo sapiens.
T36481                                                                 FH   Key                Location/Qualifiers
ID   T36481 standard; cDNA; 3415 BP.                                   FT   5'UTR              1..1088
AC   T36481;                                                           FT                      /*tag= a
DT   08-OCT-1996 (first entry)                                         FT   CDS                1089..11189
DE   Human integrin beta subunit protein, beta-5, cDNA.                FT                      /*tag= b
KW   Human; integrin beta subunit; beta-5; carcinoma; lymphoid cell;   FT                      /product= "USF2 gene product"
```

```
FT
FT   exon            /note=.1150 "contains introns"
FT                   1089..1150
FT                   /tag= c
FT   intron          /number= 1
FT                   1151..1435
FT                   /tag= d
FT   exon            /number= 1
FT                   1436..1482
FT                   /tag= e
FT                   /number= 2
FT   intron          1483..1571
FT                   /tag= f
FT                   /number= 2
FT   exon            1572..1690
FT                   /tag= g
FT                   /number= 3
FT   intron          1691..1792
FT                   /tag= h
FT                   /number= 3
FT   exon            1793..1993
FT                   /tag= i
FT                   /number= 4
FT   intron          1994..2437
FT                   /tag= j
FT                   /number= 4
FT   exon            2438..2588
FT                   /tag= k
FT                   /number= 5
FT   intron          2589..2708
FT                   /tag= l
FT                   /number= 5
FT   exon            2709..2796
FT                   /tag= m
FT                   /number= 6
FT   intron          2797..3074
FT                   /tag= n
FT                   /number= 6
FT   exon            3075..3132
FT                   /tag= o
FT                   /number= 7
FT   intron          3133..10631
FT                   /tag= p
FT                   /number= 7
FT   exon            10632..10726
FT                   /tag= q
FT                   /number= 8
FT   intron          10727..10879
FT                   /tag= r
FT                   /number= 8
FT   exon            10888..11008
FT                   /tag= s
FT                   /number= 9
FT   intron          11009..11099
FT                   /tag= t
FT                   /number= 9
FT   exon            11100..11186
FT                   /tag= u
FT                   /number= 10
FT   3'UTR           11190..11820
FT                   /tag= v
FT   polyA_signal    11723..11728
FT                   /tag= w
PN   WO9815650-A2.
PD   16-APR-1998.
PF   09-OCT-1997; E05583.
PR   09-OCT-1996; EP-202820.
PA   (VLAA-) VLAAMS INTERUNIVERSTIAIR INST BIOTECHNOG.
PI   Fryns JPCJ, Groenen PMA, Van De Ven WJM;
DR   WPI: 98-240833/21.
DR   P-PSDB; W60569.
PT   Hydronephrosis gene - useful to treat or diagnose renal diseases and
PT   disorders, e.g. vesical-ureteral reflux, pelvi-ureteral junction
PT   obstruction, multicystic renal dysplasia or renal agenesis
```

```
PS   Disclosure; Fig 6A-B; 73pp; English.
CC   This DNA encodes a USF2 gene derived from human chromosome 19. A
CC   translocation partner to this gene on chromosome 6 is the hydronephrosis
CC   gene (HNG) product. The HNG gene can be used as a starting point to
CC   design suitable compounds or techniques for the treatment of renal
CC   diseases or disorders, or nucleotide probes for diagnosing cells involved
CC   in renal diseases or disorders. A protein or a fragment encoded by HNG
CC   gene can be used as a starting point for preparing suitable antibodies
CC   for diagnosing cells involved in renal diseases and disorders. The
CC   products and method can be used to treat or diagnose renal diseases and
CC   disorders selected from vesical-ureteral reflux, uni or bilateral
CC   pelvi-ureteral junction obstruction, multicystic renal dysplasia, renal
CC   agenesis, renal aplasia, hydronephrosis, Von Mayer-Rokitansky-Kuester
CC   disorder and bifid ureter.
SQ   Sequence 11820 BP; 745 A;  1498 C;  1443 G;  683 T;

Query Match             8.1%;  Score 36.6;  DB 1;  Length 11820;
Best Local Similarity  47.2%;  Pred. No. 1.3;
Matches 111;  Conservative  0;  Mismatches 124;  Indels  0;  Gaps  0;

Qy    210 CCTCCAGGGGGACCTGCTACTTGCCCCGAGGCTTCGCCGTCACCGGCTGCACTTGTG 269
          !!!!! |  | |! !|| |||| ||  |  | |   | |!  |!| |||   | |||
Db   1267 CCTCCATTTGAGCCGGCCCCGGCCCTCGGGATCATGGGCGGCCCCCGGCTGCAGGC 1208

Qy    270 GCTCCGCCTGTGGCTCGTGGGATGTGCGCGCGAGACCACATGTCAGTGCGCGG 329
          |   | | ||  |!| || !| |||||  |  |||||
Db   1207 CCGGGGGCCGGGGCCGGGGGCCGCGGGCCACGGGCACGGGGATCTTACCTG 1148

Qy    330 GCATGGACTGGACCGGAGCCGCGCTGCTCGTGTGCAGCCCTGAGGTCGCCGCAGCGCG 389
          !|  !|!| | !|  || || | |  |! | |  || | ||| || | |    |  |  | |
Db   1147 GCGGCGGCAGCAGCGTGCCGAGGCAGCGGGATCCAGACCCGGTCCAGCATGTCCATG 1088

Qy    390 TGCACAGCCGCGGGAGGCGGCTCCAGTCCGAGGGGTTGCGGGGGAGCTGGA 444
          |!  ||||||| |!| |!!|  || | ||| |  !!| !||| |||  |   |!!
Db   1087 GGGGGGGCCGGGGGGCCGGGGGGCCCGGGGGGAGGGAGGGAGGGGGA 1033

Search completed: May 26, 2000, 03:01:51
Job time: 10587 sec
```